

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:45:45 ; Search time 96 Seconds
(without alignments)
2274.701 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTTCQVLNPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:46:45 ; Search time 21 Seconds
(without alignments)
1956.658 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTTCQVLNPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:46:45 ; Search time 21 Seconds
(without alignments)
1956.658 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTTCQVLNPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:51:20 ; Search time 15 Seconds
(without alignments)
1533.568 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTTCQVLNPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 235405 seqs, 46284737 residues

Total number of hits satisfying chosen parameters: 235405

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /SIDS5/ptodata/1/pubpaa/US08_NEW_PUB.pep1:*
2: /SIDS5/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
3: /SIDS5/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
4: /SIDS5/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
5: /SIDS5/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
6: /SIDS5/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
7: /SIDS5/ptodata/1/pubpaa/US09_NEW_PUB.pep1:*
8: /SIDS5/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
9: /SIDS5/ptodata/1/pubpaa/US10_NEW_PUB.pep1:*
10: /SIDS5/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
11: /SIDS5/ptodata/1/pubpaa/US11_NEW_PUB.pep1:*
12: /SIDS5/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:46:35 ; Search time 41 Seconds
(without alignments)
1166.335 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTCQVLNPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:45:59 ; Search time 230 Seconds
(without alignments)
1524.555 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744

Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTTCQVLNPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

RESULT 8

US-09-548-938A-10

; Sequence 10, Application US/09548938A

; Patent No. 6573086

; GENERAL INFORMATION:

; APPLICANT: EMALFARB, MARK AARON

; APPLICANT: BURLINGAME, RICHARD PAUL

; APPLICANT: OLSON, PHILIP TERRY

; APPLICANT: SINITSYN, ARKADY PANTELEIMONOVICK

; APPLICANT: PARRICHE, MARTINE

; APPLICANT: BOUSSON, JEAN CHRISTOPHE

; APPLICANT: PYNNONEN, CHRISTINE MARIE

; APPLICANT: PUNT, PETER JAN

; APPLICANT: VAN-ZEIJL, CORNELIA MARIA JOHANNA

; TITLE OF INVENTION: TRANSFORMATION SYSTEM IN THE FIELD OF FILAMENTOUS FUNGI

; FILE REFERENCE: 3123-4001

; CURRENT APPLICATION NUMBER: US/09/548,938A

; CURRENT FILING DATE: 2000-04-13

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 10

; LENGTH: 526

; TYPE: PRT

; ORGANISM: Chrysosporium lucknowense

; FEATURE:

; NAME/KEY: MOD_RES

; LOCATION: (249)

; OTHER INFORMATION: Variable amino acid

; FEATURE:

; NAME/KEY: MOD_RES

; LOCATION: (365)

; OTHER INFORMATION: Variable amino acid

US-09-548-938A-10

Query Match 61.8%; Score 1695; DB 2; Length 526;

Best Local Similarity 60.4%; Pred. No. 4e-124;

Matches 311; Conservative 68; Mismatches 112; Indels 24; Gaps 10;

```

Qy      1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWSSSTL 60
      | : ||| | : | || | | | : ||| : | | : | | | | | | | | : | : | : | : | : | : | : |
Db      18 QNACTLTAENHPSLTWSKCTSGGSCTSVQGSITIDANWRWTHRTDSATNCYEGNKWDTSY 77

Qy      61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 119
      | | : | | | : | | | | : | | | | : | | | | : | | | : | | | | | | | | | | | |
Db      78 CSDGPSCASKCCIDGADYSSTYGITTSGNSLNLKFVTKGQYSTNIGSRTYLMESDTPKYQM 137

Qy     120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
      | | | | | | : | | | | | | | | | | | | | | | : | | | | | | | | | | | | | | | |
Db     138 FQLLGNEFTFDVDVSNLGCGLNGALYFVSMDADGGMSKYSNGKAGAKYGTGYCDSQCPRD 197

Qy     180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      | | | | | : | | | | : | : | : | | | : | | | | | | | | | | : | | | | : | | |
Db     198 LKFINGEANVENWQSSNDANAGTGKYGSCCSEMDVWEANNMAAAFTPHPCXVIGQSRCE 257

Qy     240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSGA 299
      | | | | | | : | | | | | | | | : | | | | : | | | | : | | | | : | | | | : | |
Db     258 GDSCGGTYSDRYAGICDPDGCDFNSYRQGNKTFYKG--MTVDTTKKITVVVTQFLKNSA 315

Qy     300 -----INRYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFK 353
      | | : | | | | | : : | | : | : | : | | : | | | : | | | : |
Db     316 GELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQDWCDRQKAAGDVTDXQDKGGMVQMG 375

Qy     354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
      | | : | | | | | : | | : | | | | | : : | | | | | : | | | | : | : | : |
Db     376 KALAGPMVLVMSIWDHDAVNMLWLDSTWPI-DGAGKPGAERGACPTTSGVPAEVEAEAPN 434

```

Qy 414 AKVTFSNIKFGPIGST--GNPSG--GNPPGGNPPGTTTTTRRP--ATTTGSSPGPT----- 462
 : | |||:||||| | | | || | || :|| | :||: | |||
 Db 435 SNVIFSNIRFGPIGSTVSGLPDGGSGNP---NPPVSSSTPVPSSSTTSSGSSGPTGGTGV 491
 Qy 463 QSHYQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
 || |||||:|||| | | || || :|||||
 Db 492 AKHYEQCGGIGFTGPTQCESPYTCTKLNDWYSQCL 526

Title: US-10-804-785-2

RESULT 10

US-08-676-166A-3

; Sequence 3, Application US/08676166A

; Patent No. 5955270

; GENERAL INFORMATION:

; APPLICANT: Radford, Alan

; APPLICANT: Parish, John H.

; TITLE OF INVENTION: EXPLOITATION OF THE CELLULASE COMPLEX OF

; TITLE OF INVENTION: NEUROSPORA

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David A. Jackson, Esq.

; STREET: 411 Hackensack Ave, Continental Plaza, 4th

; STREET: Floor

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/676,166A

; FILING DATE: 15-JUL-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 1321-1-002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-487-5800

; TELEFAX: 201-343-1684

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 525 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: H. grisea

US-08-676-166A-3

Query Match 60.4%; Score 1658; DB 1; Length 525;

Best Local Similarity 57.5%; Pred. No. 3.1e-121;

Matches 295; Conservative 77; Mismatches 119; Indels 22; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSSVIDANWRWTHATNSSTNCYDGNTWSSTL 60

| ||:| :| || |:|:|:| || | : :|:| || | : |||| || | :||

Db 19 QQACSLTTERHPSLSWKCTAGGQCQTQVQASITLDSNWRWTHQVSGSTNCYTGKNWDTSI 78

Qy 61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 119

```

.      | | :|:|:|:| | | | | :|:|:|:| | | : | | | : | | | : | |
Db      79 CTDAKSCAQNCCVDGADYTSTYGITTNGLSLKFKVTKGQYSTNVGSRTYLMDGEDKYQT 138

Qy      120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
      | | | | | :|:| | | : | | | | | | | | | | :|:| | | | | | | | :|:| | |
Db      139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMDADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy      180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      :|:| | | :|:| | | :|:| | | :|:| | | | | | | | :|:| | | | | |
Db      199 IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE 258

Qy      240 GDGCGGTYSNRYGGTCDPDGCDWNYPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS-- 297
      | | | | | :| | | | | | | | | | :| | | | | :|:| | | | | | | |
Db      259 GDSCGGTYSNERYAGVCDPDGCDFNSYRQGNKTFYGGK--MTVDTTKKITVVVTQFLKDN 316

Qy      298 ---GAINRYYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGG--SSFSDKGGLTQFK 353
      | | | :|:| | | : | : | | : | | : | : | | | : | | | |
Db      317 GDLGEIKRFYVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

Qy      354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPQAVESQSPN 413
      | | :| | | | | :|:| : | | | | | :| : | | | | | :|:| | | | | :|:| |
Db      377 KALAGPMVLVMSIWDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy      414 AKVTFSNIKFGPIGST-----GNPSGGNPPGPNPGTTTTRRPATTTGSSPGPTQS 464
      : | | | | :|:| | | | | | | | :|:| | | | | | | | :|:| | | | |
Db      436 SNVVFNSIRFGPIGSTVAGLPGAGNGGNGGNPP---PPTTTTSSAPATTTTASAGPKAG 492

Qy      465 HYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
      : | | | | | :|:| | | | | | | | :|:| | | | |
Db      493 RWQQCGGIGFTGPTQCEEPYTCTKLNDWYSQCL 525

```

Title: US-10-804-785-2

RESULT 15

US-10-481-179-2

```

; Sequence 2, Application US/10481179
; Publication No. US20040197890A1
; GENERAL INFORMATION:
; APPLICANT: Novozymes A/S
; APPLICANT: Lange, Lene
; APPLICANT: Wu, Wenping
; APPLICANT: Aubert, Dominique
; APPLICANT: Landvik, Sara
; APPLICANT: Schnorr, Kirk
; APPLICANT: Clausen, Ib
; TITLE OF INVENTION: Polypeptides having cellobiohydrolase I activity and
; TITLE OF INVENTION: polynucleotides encoding same
; FILE REFERENCE: 10129.204-WO
; CURRENT APPLICATION NUMBER: US/10/481,179
; CURRENT FILING DATE: 2003-12-17
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Acremonium thermophilum
US-10-481-179-2

```

```

Query Match      65.6%; Score 1799; DB 4; Length 526;
Best Local Similarity 64.5%; Pred. No. 2.8e-128;
Matches 330; Conservative 57; Mismatches 107; Indels 18; Gaps 7;

```

```

Qy      1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSSVIDANWRWTHATNSSTNCYDGNTWSSTL 60
      | | | | :| | | :| | | :| | | :| | | | | | | | :|:| | | | | :|:|
Db      18 QQACTLTAENHPTLSWSKCTSGGSCTSVSGSVTIDANWRWTHQVSSSTNCYTGNEWDTSI 77

```

QY 61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 119
 Db 78 CTDGASCAAACCLDGADYSGTYGITTSGNALSLQFVTQGPYSTNIGSRTYLMASDTKYQM 137

QY 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADAGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
 Db 138 FTLLGNEFTFDVDVTGLGCGLNGALYFVSMDEDDGGLSKYSGNKAGAKYGTGYCDSQCPRD 197

QY 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
 Db 198 LKFINGEANNVGWTPSSNDKNAGLGNYSGCCSEMDVWEANSISAAAYTPHPCTTIGQTRCE 257

QY 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSGA 299
 Db 258 GDDCGGTYSTDRYAGECDPDGCDFNSYRMGNTTFYKG--MTVDTSKKFTVVVTQFLTDSS 315

QY 300 -----INRYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGGSS-FSDKGGLTQFK 353
 Db 316 GNLSEIKRFYVQNGVVIPNSNSNIAGVSGNSITQAFCDQAQKTAFGDTNVFDQKGGLAQMG 375

QY 354 KATSGGMVLVMSLWDDYYANMLWLDDSTYPTNETSSTPGAVRGSCSTSSGVPQAVESQSPN 413
 Db 376 KALAQPMVLVMSLWDDHAVNMLWLDDSTYPTN-AAGKPGAARGTCPTTSGVPADVESQAPN 434

QY 414 AKVTFSNIKFGPIGST--GNPSGGNPPGGNPPGTTTTTRRPATTTGSSP-----GPTQSH 465
 Db 435 SKVIYSNIRFGPIGSTVSGPLPGGGSNPGGSSSTTTTTTRPATSTTSSASSGPTGGGTAAH 494

QY 466 YGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 497
 Db 495 WGQCGGIGWTGPTVCASPYTCQKLNDWYYQCL 526

Title: US-10-804-785-2

RESULT 4
 S38794

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (*Humicola grisea*)
 N;Alternate names: beta-glucan cellobiohydrolase; exoglucanase
 C;Species: *Humicola grisea* var. *thermoidea*
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C;Accession: S38794; S08240; A45869

R;Radford, A.

submitted to the EMBL Data Library, June 1991

A;Reference number: S38794

A;Accession: S38794

A;Molecule type: DNA

A;Residues: 1-525 <RAD>

A;Cross-references: UNIPROT:P15828; UNIPARC:UPI000012BE0F; EMBL:X17258; NID:g2760;

PIDN:CAA35159.1; PID:g2761

A;Note: this is a revision to the sequence from reference S08240

R;de Oliveira Azevedo, M.; Radford, A.

Nucleic Acids Res. 18, 668, 1990

A;Title: Sequence of cbh-1 gene of *Humicola grisea* var. *thermoidea*.

A;Reference number: S08240; MUID:90175006; PMID:2308855

A;Accession: S08240

A;Molecule type: DNA

A;Residues: 1-299, 'H', 301-525 <DEO>

A;Cross-references: UNIPARC:UPI00001729F6; EMBL:X17258

A;Note: the authors translated the codon CAG for residue 87 as His

A;Note: this sequence has been revised in reference S38794

R;Azevedo, M.; de, O.; Felipe, M.S.S.; Astolfi-Filho, S.; Radford, A.

J. Gen. Microbiol. 136, 2569-2576, 1990

A;Title: Cloning, sequencing and homologies of the cbh-1 (exoglucanase) gene of *Humicola grisea* var. *thermoidea*.

A;Reference number: A45869; MUID:91178527; PMID:2127803

A;Accession: A45869

A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-20,'R',22-34,'K',36-86,'H',88-141,'V',143-157,'Y',159-237,'QQH',241-244,'I',246-299,'H',301-525 <AZE>
A;Cross-references: UNIPARC:UPI00001729F7; GB:M64588; GB:X17258
A;Note: this sequence has been revised. See entry S08240
C;Genetics:
A;Gene: cbh-1
A;Introns: 138/1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;494-525/Domain: fungal cellulose-binding domain homology <FCB>

```

Query Match          60.2%;  Score 1652;  DB 1;  Length 525;
Best Local Similarity 57.3%;  Pred. No. 1.8e-91;
Matches 294;  Conservative 76;  Mismatches 121;  Indels 22;  Gaps 7;

Qy      1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANNRWTHATNSSTNCYDGNWTSSSTL 60
      | |||: | || |::|| | : :||| | : ||| | || |::
Db      19 QQACSLTTERHPSLSWNKCTAGGQCQTVQASITLDSNWRWTHQVSGSTNCYTGNKWDTSI 78

Qy      61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE 119
      | |::||:||||:| | |||:|::|::| | |:: | ||: | || | : | |
Db      79 CTDAKSCAQNCCVDGADYTSTYGITNGDSLKLFVTKGQHSTNVGSRTYLMGDEKDYQT 138

Qy      120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
      | |||||:|||| | : ||||| |||||::|::| | ||||| ||||:| |||
Db      139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMDADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy      180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      :|||::|::| | |:: | | |::||| |||||::: | |||| |:: | |
Db      199 IKFINGEANIEGWGTSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE 258

Qy      240 GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS-- 297
      || |||||: || | |||||:| || | :|| | | :|||:| |||||
Db      259 GDSCGGTYSNERYAGVCDPDGCDFNSYRQGNKTFYKG--MTVDTTKKITVVVTQFLKDAN 316

Qy      298 ---GAINRYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFK 353
      | | |::| | : : | | : | : | : | | | : | | | : |
Db      317 GDLGEIKRFYVQDGKIIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

Qy      354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
      || :| |||||:|::: | |||||:| : : ||| ||: | :|||:|:::| |
Db      377 KALAGPMVLVMSIWDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy      414 AKVTFSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQS 464
      : | |||:| |||| | : |||| | || | : |||| : | ||
Db      436 SNVVFSNIRFGPIGSTVAGLPGAGNGGNGGNPP---PPTTTTSSAPATTTTASAGPKAG 492

Qy      465 HYQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
      : |||||:| || | | || :||| |
Db      493 RWQQCGGIGFTGPTQCEEPYICTKLNDWYSQCL 525

```

Title: US-10-804-785-2

RESULT 6
S42093

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - *Neurospora crassa*

C;Species: *Neurospora crassa*

C;Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C;Accession: S42093

R;Taleb, F.; Radford, A.

submitted to the EMBL Data Library, February 1994

A;Description: Cloning sequencing and homologies of the CBH-1 (exocellobiohydrolase) gene of *Neurospora crassa*.

A;Reference number: S42093

A;Accession: S42093
A;Molecule type: DNA
A;Residues: 1-516 <TAL>
A;Cross-references: UNIPROT:P38676; UNIPARC:UPI000011D714; EMBL:X77778; NID:g456657;
PIDN:CAA54815.1; PID:g456658
C;Genetics:
A;Introns: 227/3
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;485-516/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 56.9%; Score 1561; DB 2; Length 516;
Best Local Similarity 57.5%; Pred. No. 4.7e-86;
Matches 294; Conservative 62; Mismatches 129; Indels 26; Gaps 10;

```
Qy      1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSSVIDANWRWTHATNSSTNCYDGNWTSSSTL 60
      | | | | : | | | | | : | | | | : | | | | | | | | : | | | | : | |
Db     18 QQAGTLTAKRHPSLTWQKCTRGGCPTLNT-TMVLNANWRWTHATSGSTKCYTGNKWQATL 76

Qy     61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEF 120
      | | | : | | | | | | | | | | : | | | | : | | | | | | | | | |
Db     77 CPDGTKSCAANCALDGADYTGTYGITGSGWSLTQLFVTD----NVGARAYLMADDTQYQML 132

Qy    121 TLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 180
      | | | | | | | | : | | | | | | | | : | | | | | | | | | | | |
Db    133 ELLNQELWFDVDMNSIPCGLNGALYLSAMDADGGMRKYPTNKAGAKYATGYCDAQCPRL 192

Qy    181 KFINGQANVEGWEPSSNNANTGIGGHGCCSEMDIWEANSISEALTPHPCTTVGQEICEG 240
      | : | | | | | | | | | : | : | | | | | | | | | | | | : | : | |
Db    193 KYINGIANVEGWTPSTNDAN-GIGDHGCCSEMDIWEANKVSTAFTPHPCTTIEQHMCEG 251

Qy    241 DCGGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSGA- 299
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    252 DSCGGTYSDDRYGVLCADGCFNSYRMGNTTFYGEKG--TVDTSSKFTVVVTQFIKDSAG 309

Qy    300 ----INRYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFKK 354
      | : | | | | : : : | | : : | : : | | : | | | | | |
Db    310 DLAEIKAFYVQNGKVIENSQSNVDGVSGNSITQSFCKSQKTAFGDIDDFNKKGGLKQMGK 369

Qy    355 ATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSSTPGAVRGSCSTSSGVPAQVESQSPNA 414
      | : | | | | | : | | : | | | | | | | | | | | | | | : | | : | |
Db    370 ALAQAMVLVMSIWDDHAANMLWLDSTYP---VPKVPAYRGSGPTTSGVPAEVDANAPNS 426

Qy    415 KVTFSNIKFGPI-----GSTGNPSGGNPPGGNPPGTTTTTRRPATTTGSSP-GPTQSHY 466
      | | | | | | | : | : | | | | | | | : | : | : | | | | : | :
Db    427 KVAFSNIKFGHLGISPFSGGSSGTPP-SNPSSASPTSSSTAKPSSTSTASNPSGTGAHAW 485

Qy    467 GQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
      | | | | | : | | | | | | | | : | | | : | | | :
Db    486 AQCGGIGFSGPTTCPEPYTCAKDHDIIYSQCV 516
```

Title: US-10-804-785-2

RESULT 13

JE0313

exoglucanase (EC 3.2.-.-) - imperfect fungus (*Humicola grisea*)

C;Species: *Humicola grisea*

C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 09-Jul-2004

C;Accession: JE0313

R;Takashima, S.; Iikura, H.; Nakamura, A.; Hidaka, M.; Masaki, H.; Uozumi, T.

J. Biochem. 124, 717-725, 1998

A;Title: Isolation of the gene and characterization of the enzymatic properties of a major exoglucanase of *Humicola grisea* without a cellulose-binding domain.

A;Reference number: JE0313; MUID:98429588; PMID:9756616

A;Accession: JE0313

A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-451 <TAK>
A;Cross-references: UNIPROT:O93780; UNIPARC:UPI0000005E865; DDBJ:AB003105
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase

Query Match 45.2%; Score 1241.5; DB 2; Length 451;
Best Local Similarity 52.0%; Pred. No. 4.8e-67;
Matches 226; Conservative 84; Mismatches 114; Indels 11; Gaps 9;

```
Qy      1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWSSSTL 60
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      23 QQAGTITAENHPRMTWKRCSPGNCQTVQGEVVIDANWRWLH--NNGQNCYEGNKWTSQ- 79

Qy      61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQE 119
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      80 CSSATDCAQRCALDGANYQSTYGASTSGDSLTLKFVTKHEYGTNIGSRFYLMANQNKYQM 139

Qy     120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     140 FTLMNNEFAFDVDLSKVECGINSALYFVAMEEDGGMASYPNIRAGAKYGTGYCDAQCARD 199

Qy     180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQ-EIC 238
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     200 LKFIGGKANIEGWRPSTNDPNAGVGPMGACCAEIDVWESNAYAYAFTPHACGSKNRYHIC 259

Qy     239 EGDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTVQFETSG 298
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     260 ETNNCGGTYSDDRFAGYCDANGCDYNPYRMGNKDFYKGK--TVDNTRKFTTVSRFERN- 316

Qy     299 AINRYYYVQNGVTFQQPNAEL-GSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFKKAT 356
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     317 RLSQFFVQDGRKIEVPPPTWPGLPNSADITPELCAQFRVFDDRNRFETGGFDALNEAL 376

Qy     357 SGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKV 416
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     377 TIPMVLVMSIWDDHHSNMLWLDSSYPP-EKAGLPGGDRGPCPTTSGVPAEVEAQYPDAQV 435

Qy     417 TFSNIKFGPIGSTGN 431
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     436 VWSNIRFGPIGSTVN 450
```

Title: US-10-804-785-2

RESULT 8

Q8WZJ4_PENFN

ID Q8WZJ4_PENFN PRELIMINARY; PRT; 529 AA.
AC Q8WZJ4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Xylanase/cellobiohydrolase precursor (EC 3.2.1.91).
GN Name=xynA;
OS Penicillium funiculosum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=28572;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22548831; PubMed=12664153;
RA Alcocer M., Furniss C., Kroon P.A., Campbell M., Archer D.B.;
RT "Comparison of modular and non-modular xylanases as carrier proteins
RT for the efficient secretion of heterologous proteins from Penicillium
RT funiculosum."
RL Appl. Microbiol. Biotechnol. 60:726-732(2003).
RN [2]

7
✓

RP NUCLEOTIDE SEQUENCE.
 RA Furniss C.S.M., Williamson G., Kroon P.A.;
 RT "The substrate specificity and susceptibility to wheat inhibitor
 RT proteins of *Penicillium funiculosum* xylanases from a commercial enzyme
 RT preparation.";
 RL J. Sci. Food Agric. 85:574-582(2005).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 DR EMBL; AJ312295; CAC85737.1; -; Genomic_DNA.
 DR HSSP; Q09431; 1GPI.
 DR GO; GO:0005576; C:extracellular region; IEA.
 DR GO; GO:0016162; F:cellulose 1,4-beta-cellobiosidase activity; IEA.
 DR GO; GO:0030248; F:cellulose binding; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0030245; P:cellulose catabolism; IEA.
 DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
 DR GO; GO:0045493; P:xylan catabolism; IEA.
 DR InterPro; IPR000254; CBD_fun.
 DR InterPro; IPR001722; Glyco_hydro_7.
 DR Pfam; PF00734; CBM_1; 1.
 DR Pfam; PF00840; Glyco_hydro_7; 1.
 DR PRINTS; PR00734; GLHYDRLASE7.
 DR ProDom; PD001821; CBD_fungal; 1.
 DR ProDom; PD186135; Glyco_hydro_7; 1.
 DR SMART; SM00236; fCBD; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 KW Carbohydrate metabolism; Cellulose degradation; Glycosidase;
 KW Hydrolase; Polysaccharide degradation; Signal; Xylan degradation.
 FT SIGNAL 1 24 Potential.
 FT CHAIN 25 529 xylanase/cellobiohydrolase.
 SQ SEQUENCE 529 AA; 55048 MW; 95232F53577B6416 CRC64;

Query Match 63.1%; Score 1730.5; DB 2; Length 529;
 Best Local Similarity 63.1%; Pred. No. 2.9e-103;
 Matches 323; Conservative 55; Mismatches 111; Indels 23; Gaps 9;

Qy	1	QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWTSSSTL	60
Db	26	QQIGTYTAETHPSLSWSTCKSGGSCITNSGAILDANWRWVHGVTSTNCYTGNTWNTAI	85
Qy	61	CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEF	120
Db	86	CDTDAASCAQDCALDGADYSPTYGITTSNSLRNLNFVTGS---NVGSRTYLMADNTHYQIF	142
Qy	121	TLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL	180
Db	143	DLNLQEFTFITVDVSNLPCGLNGALYFVTMDADGGVSKYPNNKAGAQQYGVGYCDSQCPRDL	202
Qy	181	KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG	240
Db	203	KFIAGQANVEGWTPSTNNSNTGIGNHGSCCAELDIWEANSISEALTPHPCDTPGLTVCTA	262
Qy	241	DGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVQTQFET----	296
Db	263	DDCGGTYSSNRYAGTCDPDGCDFNPYRLGVTDIFYGSGK--TVDTTKPFTVVVTQFVTDDGT	320
Qy	297	-SGA---INRYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFG-GSSFSDKGGTLTQ	351
Db	321	SSGSLSEIRRYVQNGVVIPQPSSKISGISGNVINSDFCAAELSAFGETASFTNHGGLKN	380
Qy	352	FKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQS	411
Db	381	MGSALEAGMVLVMSLWDDYSVNMLWLDSTYPANET-GTPGAARGSCPTTSGNPKTVESQS	439
Qy	412	PNAKVTFSNIKFGPIGSTGNPSGGNPPGGN----PPGTTTTRRPATTTGSSPGPT--QSH	465

Db 440 GSSYVVFSDIKVGPFNSTF--SGGTSTGGSTTTTASGTTSTKASTTSTSSTSTGTGVAAH 497

Qy 466 YGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
:||||| |:| ||| ||||| |:|||||||

Db 498 WGQCGGQGWGTGPTTCASGTTCTVVPYYSQCL 529

Title: US-10-804-785-2

RESULT 14

Q12621_HUMGT

ID Q12621_HUMGT PRELIMINARY; . PRT; 525 AA.
AC Q12621;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cellulase (EC 3.2.1.91).
GN Name=cbh-1;
OS Humicola grisea var. thermoidea.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=5528;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IFO9854;
RA Takashima S., Nakamura A., Hidaka M., Masaki H., Uozumi T.;
RT "Cloning, sequencing, and expression of the cellulase genes of
RT Humicola grisea var. thermoidea."
RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: The biological conversion of cellulose to glucose
CC generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC Exocellobiohydrolases that cut the dissaccharide cellobiose from
CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
CC glucosidases which hydrolyze the cellobiose and other short cello-
CC oligosaccharides to glucose (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
DR EMBL; D63515; BAA09785.1; -; Genomic_DNA.
DR HSSP; Q09431; 1GPI.
DR GO; GO:0005576; C:extracellular region; IEA.
DR GO; GO:0016162; F:cellulose 1,4-beta-cellobiosidase activity; IEA.
DR GO; GO:0030248; F:cellulose binding; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030245; P:cellulose catabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR000254; CBD_fun.
DR InterPro; IPR001722; Glyco_hydro_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR PRINTS; PR00734; GLHYDRLASE7.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; Glyco_hydro_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Carbohydrate metabolism; Cellulose degradation; Glycosidase;
KW Hydrolase; Polysaccharide degradation.
SQ SEQUENCE 525 AA; 55722 MW; A2E6E5F40F6D3BB0 CRC64;

Query Match 60.4%; Score 1658; DB 2; Length 525;
Best Local Similarity 57.5%; Pred. No. 1.3e-98;
Matches 295; Conservative 77; Mismatches 119; Indels 22; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSSVVIDANNRWTHATNSSTNCYDGNTWSSTL 60
| ||:| :| || |:|:|:|:| | | : :|:| || | :|:|

Db 19 QQACSLTTERHPSLSWKKCTAGGQCQTQVQASITLDSNWRWTHQVSGSTNCYTGNKWDTSI 78

```

Qy      61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 119
      | | :|:|:|:|:| | | | | :|:|:|:|:| :| | :| | | :| |
Db      79 CTDAKSCAQNCQCCVDGADYTSTYGITNGDSLKFKVTKGQYSTNVGSRITYLMDGEDKYQT 138

Qy     120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
      | | | | | :| | | | : | | | | | | | | | | :|:| | | | | | | | :| | | |
Db     139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy     180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      :| | | | :|:| | | :|:| | | | | | | | | | :| | | | :| | |
Db     199 IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE 258

Qy     240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS-- 297
      | | | | | :| | | | | | | | | | | | | | | | | | | | | | | |
Db     259 GDSCGGTYSNERYAGVCDPDGCDFNSYRQGNKTFYGKG--MTVDTTKKITVVVTQFLKDN 316

Qy     298 ---GAINRYYYQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFK 353
      | | | :| | | :| | :| | :| | :| | :| | :| | :| | :| |
Db     317 GDLGEIKRFYVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

Qy     354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPQAVESQSPN 413
      | | :| | | | | :| | :| | | | | :| | :| | | | | :| | | | | :| |
Db     377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy     414 AKVTFSNIKFGPIGST-----GNPSGGNPPGPNPGTTTTTRRPATTTGSSPGPTQS 464
      :| | | | :| | | | | | | | | | | | | | | | | | | | | |
Db     436 SNVVFNSIRFGPIGSTVAGLPGAGNGGNGGNPP---PPTTTTSSAPATTTTASAGPKAG 492

Qy     465 HYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
      :| | | | :| | | | | | | | | | | | | | | | | |
Db     493 RWQQCGGIGFTGPTQCEEPYTCTKLNDWYSQCL 525

```

Title: US-10-804-785-2

RESULT 15

GUX1_HUMGT

```

ID  GUX1_HUMGT      STANDARD;      PRT;      525 AA.
AC  P15828;
DT  01-APR-1990 (Rel. 14, Created)
DT  01-FEB-1996 (Rel. 33, Last sequence update)
DT  10-MAY-2005 (Rel. 47, Last annotation update)
DE  Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (1,4-
DE  beta-cellobiohydrolase) (Beta-glucancellobiohydrolase).
GN  Name=CBH-1;
OS  Humicola grisea var. thermoidea.
OC  Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX  NCBI_TaxID=5528;
RN  [1]
RP  NUCLEOTIDE SEQUENCE.
RX  MEDLINE=90175006; PubMed=2308855;
RA  de Oliviera Alzevedo M., Radford A.;
RT  "Sequence of cbh-1 gene of Humicola grisea var. thermoidea.";
RL  Nucleic Acids Res. 18:668-668(1990).
CC  -!- FUNCTION: The biological conversion of cellulose to glucose
CC  generally requires three types of hydrolytic enzymes: (1)
CC  Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC  Exocellobiohydrolases that cut the dissaccharide cellobiose from
CC  the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
CC  glucosidases which hydrolyze the cellobiose and other short cello-
CC  oligosaccharides to glucose.
CC  -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC  in cellulose and cellotetraose, releasing cellobiose from the non-
CC  reducing ends of the chains.
CC  -!- SIMILARITY: Belongs to the glycosyl hydrolase 7 (cellulase C)
CC  family.
CC  -!- SIMILARITY: Contains 1 CBM1 (fungal-type carbohydrate-binding)

```

```
CC      ,      domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; X17258; CAA35159.1; -; Genomic_DNA.
DR PIR; S38794; S38794.
DR HSSP; Q09431; 1GPI.
DR InterPro; IPR000254; CBD_fun.
DR InterPro; IPR001722; Glyco_hydro_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR PRINTS; PR00734; GLHYDRLASE7.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; Glyco_hydro_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Carbohydrate metabolism; Cellulose degradation; Glycoprotein;
KW Glycosidase; Hydrolase; Polysaccharide degradation; Signal.
FT SIGNAL      1      18      Potential.
FT CHAIN       19     525     Exoglucanase I.
FT DOMAIN     490     525     CBM1.
FT REGION     19     467     Catalytic.
FT REGION    468     489     Linker.
FT ACT_SITE   231     231     Nucleophile (By similarity).
FT ACT_SITE   236     236     Proton donor (By similarity).
FT CARBOHYD   289     289     N-linked (GlcNAc . . .) (Potential).
FT DISULFID   497     514     By similarity.
FT DISULFID   508     524     By similarity.
SQ SEQUENCE   525 AA;  55694 MW;  A6684D4CF881E090 CRC64;
```

Query Match 60.2%; Score 1652; DB 1; Length 525;
Best Local Similarity 57.3%; Pred. No. 3.3e-98;
Matches 294; Conservative 76; Mismatches 121; Indels 22; Gaps 7;

```
Qy      1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWSSSTL 60
      | ||:| :| || |:| ||::|| | : :||| || | :||| || | ::
Db      19 QQACSLTTERHPSLSWNKCTAGGQCQTVQASITLDSNWRWTHQVSGSTNCYTGKNKWDTSI 78

Qy     61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE 119
      | | ::||:||||:|| | ||||:||||:||||: |||: |||: ||| : ||
Db     79 CTDAKSCAQNCCVDGADYTSTYGITTTNGDSLKLFVTKGQHSTNVGSRTYLMGDEDKYQT 138

Qy    120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
      | |||||:||||| : ||||| ||||| |||||:|:| | ||||| |||||:|||||
Db    139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMDADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy    180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      :||| |:|:| | |:|: | | |:|:||||| |||||:|:| | ||||| :| | ||
Db    199 IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE 258

Qy    240 GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS-- 297
      || |||||: || | |||||:| || || :|| | |:|||:|||||
Db    259 GDSCGGTYSNERYAGVCDPDGCDFNSYRQGNKTFYGKG--MTVDTTKKITVVVTQFLKDAN 316

Qy    298 ---GAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFK 353
      | | |:| |:| : : || : |:| :: || |: |||: |
Db    317 GDLGEIKRFYVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGMKQMG 376

Qy    354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSSTPGAVRGSCSTSSGVPAQVESQSPN 413
      || :| |||||:| |:| :||| |:| :| ||| |:| |:| |:| |:| ::||
Db    377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy    414 AKVTFSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQS 464
```

Db : | | | | : | | | | | | | | | | : | | | | | | | | : | | | | : | | | |
436 SNVVFSNIRFGPIGSTVAGLPGAGNGGNNGGNPP---PPTTTTSSAPATTTTASAGPKAG 492

Qy 465 HYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497

Db : | | | | : | | | | | | | | | | : | | | | | | | | : | | | | | | | |
493 RWQQCGGIGFTGPTQCEEPYICTKLNDWYSQCL 525

Db 138 FTLLGNEFTFDVDVTGLGCGLNALYFVSMDEDGGLSKYSGNKAGAKYGTGYCDSQCPRD 197

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
 |||||:| | ||||: | |:| :|||||||:||||||| | |||||:| | ||

Db 198 LKFINGEANNVGWTPSSNDKNAGLGNYGSCCSEMDVWEANSISAAAYTPHPCTTIGQTRCE 257

Qy 240 GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTQFETSGA 299
 || ||||| :| | |||||:| ||:|:| | | :|:| | ||||| | :

Db 258 GDDCGGTYSTDRYAGECDPDGCDFNSYRMGNTTFYKGK--MTVDTSKKFTVVTQFLTDSS 315

Qy 300 -----INRYVQNGVTFFQQPNAELGSYSGNELNDYCTAEAEFGGSS-FSDKGGLTQFK 353
 | |:||||| | : : ||| : :| |:| || :| ||||| |

Db 316 GNLSEIKRFYVQNGVVIPNSNSNIAGVSGNSITQAFCDQAQKTAFGDTNVFDQKGGLAQMG 375

Qy 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
 || : |||||||||: ||||||||| : ||| ||:| |:||||| ||||:|

Db 376 KALAQPMVLVMSLWDDHAVNMLWLDSTYPTN-AAGKPGAARGTCPTTSGVPADVESQAPN 434

Qy 414 AKVTFSNIKFGPIGST--GNPSGGNPPGGNPPGTTTTTRRPATTTGSSP-----GPTQSH 465
 :|| :||:||||| | | |: || | ||| ||||:| || | | :|

Db 435 SKVIYSNIRFGPIGSTVSGLPGGGSGNPGGSSSTTTTTTRPATSTTSSASSGPTGGGTAH 494

Qy 466 YGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 497
 :||||||:|:|||||| ||| || :| |||

Db 495 WGQCGGIGWTGPTVCASPYTCQKLNDWYYQCL 526

RESULT 134

AAB81926

ID AAB81926 standard; protein; 529 AA.

XX

AC AAB81926;

XX

DT 25-JUN-2001 (first entry)

XX

DE Acremonium cellulolyticus cellobiohydrolase 1 precursor.

XX

KW Cellobiohydrolase 1; cbh1; promoter; protein production.

XX

OS Acremonium cellulolyticus.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	Peptide	1. .26
----	---------	--------

FT		/label= signal_peptide
----	--	------------------------

FT	Protein	27. .529
----	---------	----------

FT		/label= mature_cellobiohydrolase
----	--	----------------------------------

XX

PN JP2001017180-A.

XX

PD 23-JAN-2001.

XX

PF 06-JUL-1999; 99JP-00191221.

XX

PR 06-JUL-1999; 99JP-00191221.

XX

PA (MEIJ) MEIJI SEIKA KAISHA LTD.

PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX

DR WPI; 2001-294133/31.

DR N-PSDB; AAF85588.

XX

PT New promotor useful for expression of a protein.

XX

PS Disclosure; Page 12-14; 22pp; Japanese.

XX

CC The present invention provides a promoter capable of causing the

CC expression of a gene connected downstream. It can be used for expressing

FT /note= "Encoded by ACC"
FT Misc-difference 365
FT /note= "Encoded by TTN"
FT Binding-site 496. .526
FT /label= Cellulose_binding_domain
XX
PN WO200179507-A2.
XX
PD 25-OCT-2001.
XX
PF 17-APR-2001; 2001WO-NL000301.
XX
PR 13-APR-2000; 2000EP-00201343.
XX
PA (EMAL/) EMALFARB M A.
XX
PI Emalfarb MA, Punt PJ, Van Zeijl CMJ;
XX
DR WPI; 2002-066369/09.
DR N-PSDB; AAI72045.
XX
PT New glycosyl hydrolase family 7, glycosyl hydrolase family 10 and
PT glyceraldehyde phosphate dehydrogenase genes from the filamentous fungus
PT Chrysosporium useful for the microbial production of these proteins.
XX
PS Claim 1; Page 34; 43pp; English.
XX
CC This sequence shows a Chrysosporium glycosyl hydrolase family 7 protein,
CC CBH1. The CBH1 nucleic acid is used for the industrial production of CBH1
CC protein by microbial fermentation. The CBH1 regulatory sequences
CC (promoter and terminator) are useful for expressing heterologous
CC polypeptides in microbes
XX
SQ Sequence 526 AA;

Query Match 61.6%; Score 1689; DB 5; Length 526;
Best Local Similarity 60.4%; Pred. No. 5.3e-101;
Matches 311; Conservative 68; Mismatches 112; Indels 24; Gaps 10;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSSVIDANWRWTHATNSSTNCYDGNWTSSSTL 60
|:|||| |:| ||| ||:||||:| ||: ||||| ||:|:||||:| |::
Db 18 QNACTLTAENHPSLTWSKCTSGGSCTSVQGSITIDANWRWTHRTDSATNCYEGNKWDTSY 77

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 119
| | :|| ||:||| |:||||:|||||||:: ||: | |:| ||| ||| ||
Db 78 CSDGPSCASKCCIDGADYSSTYGITTSGNSLNLKFVTKGQYSTNIGSRTYLMESDTKYQM 137

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
| |||||:||||| | ||||| |||||:||| | ||||| ||||| |||||
Db 138 FQLLGNEFTFDVDVSNLGCGLNGALYFVSMDADGGMSKYSKNKAGAKYGTGYCDSQCPRD 197

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
|||||:|||| |: |:|:| | | :|||||:||||:| | |||| |:| ||
Db 198 LKFINGEANVENWQSSTNDANAGTGKYGSCCSEMDVWEANNMAAAFTPHPCWVIGQSRCE 257

Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTTQFETSGA 299
|| ||||| |:| | |||||:| || | :||| | |:||||:||||| :|
Db 258 GDSCGGTYSTDRYAGICDPDGCDFNSYRQGNKTFYKG--MTVDTTKKITVVTQFLKNSA 315

Qy 300 -----INRYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGGSSF-SDKGGLTQFK 353
| |:||||| : : || : |:| ::| | : ||||:|
Db 316 GELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQDWCDRQKAAGDVTDXQDKGGMVQMG 375

Qy 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPQAVESQSPN 413
|| :| |||||:|||| | |||||:| : : ||| ||:| |:||||:|:|:|
Db 376 KALAGPMVLVMSIWDHDAVNMLWLDSTWPI-DGAGKPGAERGACPTTSGVPAEVEAEAPN 434

```

Qy      414 AKVTFSNIKFGPIGST--GNPSG--GNPPGGNPPGTTTTTRRP--ATTGSSPGPT----- 462
      : | ||||:||||| | | | || | || :::| | :||: | ||
Db      435 SNVIFSNIRFGPIGSTVSGLPDGGSGNP--NPPVSSSTPVPSSSTTSSGSSGPTGGTGV 491

Qy      463 QSHYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
      || |||||:|||| | | || || :||||
Db      492 AKHYEQCGGIGFTGPTQCESPYTCTKLNDWYSQCL 526

```

RESULT 138

ABJ26886

ID ABJ26886 standard; protein; 529 AA.

XX

AC ABJ26886;

XX

DT 08-MAY-2003 (first entry)

XX

DE Cellobiohydrolase I activity protein SEQ ID No 4.

XX

KW Cellobiohydrolase; enzyme; DNA shuffling; ethanol; biomass;

KW cellobiohydrolase I; EC 3.2.1.91.

XX

OS Chaetomium thermophilum.

XX

PN WO2003000941-A2.

XX

PD 03-JAN-2003.

XX

PF 26-JUN-2002; 2002WO-DK000429.

XX

PR 26-JUN-2001; 2001DK-00001000.

XX

PA (NOVO) NOVOZYMES AS.

XX

PI Lange L, Wu W, Aubert D, Landvik S, Schnorr KM, Clausen IG;

XX

DR WPI; 2003-278244/27.

DR N-PSDB; ABT23504.

XX

PT New polypeptide with cellobiohydrolase I activity, useful in producing

PT ethanol from biomass.

XX

PS Claim 4; Page 115-117; 199pp; English.

XX

CC The invention relates to a novel polypeptide comprising: part of any of
CC 21 amino acid sequences; an amino acid sequence at least 70% identical to
CC a polypeptide encoded by a cellobiohydrolase gene; an amino acid sequence
CC at least 80% identical to the polypeptide encoded by 21 nucleotide
CC sequences; a polypeptide encoded by a nucleotide sequence which
CC hybridises with a probe selected from complementary strands of 55
CC nucleotide sequences; or a fragment of the aforementioned structures. The
CC polynucleotides of the invention are useful in a method of DNA shuffling.
CC The polypeptides are useful in a method for producing ethanol from
CC biomass comprising contacting the biomass with the polypeptides. This
CC sequence represents a protein with cellobiohydrolase I activity of the
CC invention

XX

SQ Sequence 529 AA;

Query Match 61.6%; Score 1689; DB 6; Length 529;

Best Local Similarity 59.3%; Pred. No. 5.3e-101;

Matches 305; Conservative 76; Mismatches 113; Indels 20; Gaps 7;

```

Qy      1 QSACTLQSETHPPLTWQKSSGGTCTQQTGSSVIDANNRWTHATNSSTNCYDGNWSSSTL 60
      | ||:| :|||| ||||:|:|||| |: |:| ||||| || | :||| || || :::
Db      19 QQACSLTTETHPRLTWKRCTSGGNCSTVNGAVTIDANNRWTHTVSGSTNCYTGNWDTSI 78

```

Qy	61	CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQE	119
Db	79	CSDGKSCAQTCCTVDGADYSSYGITTSGLSLNLKFVTKHQYGTNVGSRVYLMENDTKYQM	138
Qy	120	FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD	179
Db	139	FELLGNEFTFDVDVSNLGCGLNGALYFVSMADGGMSKYSNGKAGAKYGTGYCDAQCPRD	198
Qy	180	LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE	239
Db	199	LKFINGEANIENWTPSTNDANAGFGRYGSCCSEMDIWEANNMATAFTPHPCTIIGQSRCE	258
Qy	240	GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSGA	299
Db	259	GNSCGGTYSSERYAGVCDPDGCDFNAYRQDKTFYKG--MTVDTTKKMTVVVTQFHKNSA	316
Qy	300	-----INRYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFK	353
Db	317	GVLSEIKRFYVQDGKVIANAESKIPGNPGNSITQEWCDQKVAFGDIDDFNRKGGMAQMS	376
Qy	354	KATSGGMVLMVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPQVESQSPN	413
Db	377	KALEGPMVLMVMSVWDDHYANMLWLDSTYPIDK-AGTPGAERGACPTTSGVPAEIEAQVPN	435
Qy	414	AKVTFSNIKFGPIGSTGNPSGGNPPGG-----NPPGTTTTTRRPATTTGSSP-----GPTQ	463
Db	436	SNVIFSNIRFGPIGSTVPGLDGSTPSNPTATVAPPTSTTSVRSSTTQISTPTSQPGGCTT	495
Qy	464	SHYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL	497
Db	496	QKWGQCGGIGYTGCTNCVAGTTCTELNPWYSQCL	529

RESULT 140
 ABB05058
 ID ABB05058 standard; protein; 526 AA.
 XX
 AC ABB05058;
 XX
 DT 11-SEP-2003 (revised)
 DT 27-MAR-2002 (first entry)
 XX
 DE Trichoderma reesei cellobiohydrolase I (CBH1) 55kD (family 7) protein.
 XX
 KW Trichoderma reesei; filamentous fungi; phenotype; characterisation;
 KW fermentation; screening; morphology; cellobiohydrolase I; CBH1.
 XX
 OS Hypocrea jecorina.
 XX

FH	Key	Location/Qualifiers
FT	Peptide	1. .19
FT		/label= signal
FT	Protein	20. .526
FT		/label= cellobiohydrolase_I
FT	Misc-difference	249
FT		/label= unknown
FT		/note= "encoded by NCC"
FT	Misc-difference	365
FT		/label= unknown
FT		/note= "encoded by TTN"

 XX
 PN WO200125468-A1.
 XX
 PD 12-APR-2001.
 XX
 PF 13-APR-2000; 2000WO-US010199.

XX
PR 06-OCT-1999; 99WO-NL000618.
XX
PA (EMAL/) EMALFARB M A.
XX
PI Emalfarb MA;
XX
DR WPI; 2001-281733/29.
DR N-PSDB; ABA92722.
XX
PT Expressing heterologous proteins encoded by a library of DNA vectors,
PT involves stably transforming mutant filamentous fungus with the vectors
PT and culturing transformed fungi for expressing heterologous proteins.
XX
PS Disclosure; Page 66-69; 85pp; English.
XX
CC The present invention describes a method of expressing a number of
CC proteins encoded by a library of DNA vectors (I). The method involves
CC stably transforming a mutant filamentous fungus (II) with (I) so as to
CC introduce into each of a number of individual fungi, at least one
CC heterologous protein-encoding nucleic acid sequence (III), and culturing
CC the transformed mutant filamentous fungi for the expression of
CC heterologous proteins encoded by (III). (I) comprises a number of
CC different vectors, each comprising a different protein-encoding nucleic
CC acid sequence being operably linked to an expression regulating region
CC and optionally a secretion signal encoding sequence. (II) has a phenotype
CC characterised by a culture viscosity, when cultured in suspension, of
CC less than 200 cP at the end of fermentation when grown with adequate
CC nutrients under optimal or near-optimal conditions. The method is useful
CC for expressing large quantities of heterologous proteins that are useful
CC for isolation, characterisation and application testing, and also for
CC commercial production of proteins. The mutant filamentous fungi obtained
CC by the method are suitable for high-throughput screening techniques owing
CC to their unique morphology and very low viscosity of their cultures. The
CC present sequence represents the Trichoderma reesei cellobiohydrolase I
CC (CBH1) 55kD (family 7) protein, which is given in the exemplification of
CC the present invention. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 526 AA;

Query Match 61.4%; Score 1681; DB 4; Length 526;
Best Local Similarity 60.4%; Pred. No. 1.7e-100;
Matches 311; Conservative 68; Mismatches 112; Indels 24; Gaps 10;

Qy 1 QSACTLQSETHPPLTWQKSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWTWSSTL 60
|:|||| |:| || ||| |:|:|:| ||: ||| ||| |:|:| |||:| | ::
Db 18 QNACTLTAENHPSLTWSKCTSGGSCTSVQGSITIDANWRWTHRTDSATNCYEGNKWDTSY 77

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 119
| | :|| ||:|| |:| |||:| ||| |||:|:| ||: |:|:| ||| ||| ||
Db 78 CSDGPSCASKCCIDGADYSSTYGITTSNLSNLKFVTKGQYSTNIGSRTYLMESDTPKYQM 137

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
| |||||:| ||| | ||||| ||||| |||||:| | ||||| ||||| |||||
Db 138 FQLLGNEFTFDVDVSNLGCGLNGALYFVSMDADGGMSKYSNGKAGAKYGTGYCDSQCPRD 197

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
||| |||:| || |: |:|:| | | :|| |||:| |||:|:| | ||| |:| ||
Db 198 LKFINGEANVENWQSSTNDANAGTGKYGCCSEMDVWEANNMAAAFTPHPCXVIGQSRCE 257

Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSGA 299
|| ||||| |:| | |||||:| || || |:| | |:| |||:| ||||| |:|
Db 258 GDSCGGTYSTDYAGICDPDGCDNFNSYRQGNKTFYKGK--MTVDTTKKITVVVTQFLKNSA 315

Qy 300 -----INRYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGGSSF-SDKGGLTQFK 353
| |:| ||| |:| :| || |:| |:| |:| |:| |:| |:| |:|
Db 316 GELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQDWCDRQKAAGDVTDXQDKGGMVQMG 375

QY	354	KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN	413
		: : : : : : : : : : :	
Db	376	KALAGPMVLVMSIWDDHAVNMLWLDSTWPI-DGAGKPGAERGACPTTSGVPAEVEAEAPN	434
QY	414	AKVTFSNIKFGPIGST--GNPSG--GNPPGGNPPGTTTTTRRP--ATTTGSSPGPT-----	462
		: : : : : :	
Db	435	SNVIFSNIRFGPIGSTVSGLPDGGSGNP---NPPVSSSTPVPSSSTTSSGSSGPTGGTGV	491
QY	463	QSHYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL	497
		: : :	
Db	492	AKHYEQCGGIGFTGPTQCESPYTCTKLNDWYSQCL	526

RESULT 141

ABW00703

ID ABW00703 standard; protein; 526 AA.

XX

AC ABW00703;

XX

DT 15-JAN-2004 (first entry)

XX

DE Chrysosporium lucknowense cellobiohydrolase (CBH1) protein.

XX

KW Mutant Chrysosporium strain; fungal enzyme; metabolite; organic acid;

KW antibiotic; cellobiohydrolase; CBH1.

XX

OS Chrysosporium lucknowense.

XX

FH	Key	Location/Qualifiers
FT	Peptide	1. .20
FT		/label= Signal_peptide
FT	Protein	21. .526
FT		/note= "Mature CBH1 protein"
FT	Misc-difference	137
FT		/note= "Encoded by AGTAAGTTCCTCTCGCACCCGCCGCCGGGAGATGAT
FT		GGCGCCCAGCCCGCTGACGCGAATGACACAGTG"
FT	Misc-difference	249
FT		/note= "Encoded by ACC"
FT	Misc-difference	365
FT		/note= "Encoded by TTN"
FT	Domain	495. .526
FT		/note= "CBD domain"

XX

PN US6573086-B1.

XX

PD 03-JUN-2003.

XX

PF 13-APR-2000; 2000US-00548938.

XX

PR 06-OCT-1998; 98WO-EP006496.

PR 06-OCT-1999; 99WO-NL000618.

XX

PA (DYAD-) DYADIC INT INC.

XX

PI Emalfrab MA, Burlingame RP, Olson PT, Sinitsyn AP, Parriche M;

PI Bousson JC, Pynnonen CM, Punt PJ, Van Zeijl CMJ;

XX

DR WPI; 2003-764575/72.

DR N-PSDB; AAD61474.

XX

PT New mutant Chrysosporium strain expressing a heterologous polypeptide, or

PT overexpressing a homologous polypeptide, at a high level, useful for

PT production of e.g. enzymes, primary metabolites, and antibiotics.

XX

PS Disclosure; Col 43-44; Opp; English.

XX

CC The invention relates to a mutant Chrysosporium strain comprising a
CC nucleic acid encoding a polypeptide of interest, linked to an expression
CC -regulating region chosen from promoter sequences associated with
CC cellulase, xylanase or glyceraldehyde-3-phosphate dehydrogenase (gpdA)
CC expression and optionally to a secretion signal sequence, where the
CC mutant strain expresses the polypeptide at a higher level than a non-
CC mutant strain under same conditions. The invention is useful for
CC producing polypeptides such as carbohydrate-degrading enzymes, proteases,
CC lipases, esterases, other hydrolases, oxidoreductases and transferases.
CC The invention is also useful for producing fungal enzymes allowing
CC production or overproduction of primary metabolites, organic acids,
CC secondary metabolites or antibiotics. The present sequence is
CC Chrysosporium lucknowense cellobiohydrolase (CBH1) protein
XX
SQ Sequence 526 AA;

Query Match 61.4%; Score 1681; DB 7; Length 526;
Best Local Similarity 60.4%; Pred. No. 1.7e-100;
Matches 311; Conservative 68; Mismatches 112; Indels 24; Gaps 10;

QY 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSSVIDANWRWTHATNSSTNCYDGNWSSSTL 60
|:|||| |:| || ||| ||:||||| ||: ||||| ||:|:||||:| |::
Db 18 QNACTLTAENHPSLTWSKCTSGGSCTSVQGSITIDANWRWTHRTDSATNCYEGNKWDTSY 77

QY 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTTYQE 119
| | :|| ||:||| |:||||:|||||||:: |||: | |:| ||| ||| ||
Db 78 CSDGPSCASKCCIDGADYSSTYGITTSNGNSLNLKFVTKGQYSTNIGSRTYLMESDTKYQM 137

QY 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
| |||||:||||| | |||||:|||||:|||| | |||||:|||||
Db 138 FQLLGNEFTFDVDVSNLGCGLNGALYFVSMDADGGMSKYSNGKAGAKYGTGYCDSQCPRD 197

QY 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTTPHPCCTTVGQEICE 239
|||||:|||| |: |:|:| | | :|||||:||||:|:| ||||| :|| ||
Db 198 LKFINGEANVENWQSSSTNDANAGTGKYGSCCSEMDVWEANNMAAAFTPHPCXVIGQSRCE 257

QY 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSGA 299
|| ||||| :|| | |||||:| || || :||| | |:||||:||||| :|
Db 258 GDSCGGTYSDRYAGICDPDGCDFNSYRQGNKTFYKG--MTVDTTKKITVVVTQFLKNSA 315

QY 300 -----INRYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGSSSF-SDKGGLTQFK 353
| |:||||| : : || : |:| ::| || : ||||: |
Db 316 GELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQDWCDRQKAAFQGDVTDXQDKGGMVQMG 375

QY 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
|| :| |||||:||||: |||||:| : : ||| ||:| |:||||:|:|:|
Db 376 KALAGPMVLVMSIWDHVNMLWLDSTWPI-DGAGKPGAERGACPTTSGVPAEVEAEAPN 434

QY 414 AKVTFSNIKFGPIGST--GNPSG--GNPPGGNPPGTTTTTRP--ATTTGSSPGPT----- 462
: | ||||:||||| | | | ||| ||:| | :||: | |||
Db 435 SNVIFSNIKFGPIGSTVSGLPDGGSGNP---NPPVSSSTPVPSSSTTSSGSSGPTGGTGV 491

QY 463 QSHYQCGGIGYSGPTVCASGTTTCQVLNPPYYSQCL 497
|| |||||:|||| | | || | :|||||
Db 492 AKHYEQCGGIGFTGPTQCESPYTCTKLNDWYSQCL 526

RESULT 142
ABJ26904
ID ABJ26904 standard; protein; 525 AA.
XX
AC ABJ26904;
XX
DT 08-MAY-2003 (first entry)
XX
DE Cellobiohydrolase I activity protein SEQ ID No 60.
XX

KW Cellobiohydrolase; enzyme; DNA shuffling; ethanol; biomass;
KW cellobiohydrolase I; EC 3.2.1.91.
XX
OS Scytalidium thermophilum.
XX
PN WO2003000941-A2.
XX
PD 03-JAN-2003.
XX
PF 26-JUN-2002; 2002WO-DK000429.
XX
PR 26-JUN-2001; 2001DK-00001000.
XX
PA (NOVO) NOVOZYMES AS.
XX
PI Lange L, Wu W, Aubert D, Landvik S, Schnorr KM, Clausen IG;
XX
DR WPI; 2003-278244/27.
DR N-PSDB; ABT23542.
XX
PT New polypeptide with cellobiohydrolase I activity, useful in producing
PT ethanol from biomass.
XX
PS Claim 4; Page 191-192; 199pp; English.
XX
CC The invention relates to a novel polypeptide comprising: part of any of
CC 21 amino acid sequences; an amino acid sequence at least 70% identical to
CC a polypeptide encoded by a cellobiohydrolase gene; an amino acid sequence
CC at least 80% identical to the polypeptide encoded by 21 nucleotide
CC sequences; a polypeptide encoded by a nucleotide sequence which
CC hybridises with a probe selected from complementary strands of 55
CC nucleotide sequences; or a fragment of the aforementioned structures. The
CC polynucleotides of the invention are useful in a method of DNA shuffling.
CC The polypeptides are useful in a method for producing ethanol from
CC biomass comprising contacting the biomass with the polypeptides. This
CC sequence represents a protein with cellobiohydrolase I activity of the
CC invention
XX
SQ Sequence 525 AA;

Query Match 60.8%; Score 1666; DB 6; Length 525;
Best Local Similarity 57.5%; Pred. No. 1.6e-99;
Matches 295; Conservative 78; Mismatches 118; Indels 22; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWSSSTL 60
| |||:| :| || |:|:|:|:| | | :| :|:| | | | :| | | | | | :| :|
Db 19 QQACSLTTERHPSLSWKKCTAGGQCQTVQASITLDSNWRWTHQVSGSTNCYTGKNKWDTSI 78

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE 119
| | :|:|:|:|:| | | | |:|:|:|:|:| | | :| | | | | :| |
Db 79 CTDAKSQAQNCVDGADYTSTYGITTNGDSLKFKVTKGQHSTNVGSRTYLMDGEDKYQT 138

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
| | | | | | :| | | | | :| | | | | | | | | | | :| :| | | | | | | :| | | |
Db 139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
:| | | |:|:| | | :|:| | | | :|:| | | | | | |:|:| | | | | | :| | |
Db 199 IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE 258

Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS-- 297
| | | | | | :| | | | | | | | | | | :| | | | | | | | | | | | | | | |
Db 259 GDSCGGTYSNERYAGVCDPDGCDFNSYRQGNKTFYKGK--MTVDTTKKITVVVTQFLKDAN 316

Qy 298 ---GAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGGTLQFK 353
| :|:|:|:| :| :| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 317 GDLGEVKRFYVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

Qy 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
 || :| |||||:||||: |||||:| : : ||| ||: |:|||||:|::||
 Db 377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy 414 AKVTFSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQS 464
 : | ||||:||||| | :|||| || |||: |||| :| ||
 Db 436 SNVVFNSIRFGPIGSTVAGLPGAGNGGNNGGNPP---PPTTTTSSAPATTTTASAGPKAG 492

Qy 465 HYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
 |: |||||:|||| | || || :|||||
 Db 493 HWQQCGGIGFTGPTQCEPYTCTKLNDWYSQCL 525

RESULT 9

US-09-548-938A-10

; Sequence 10, Application US/09548938A
 ; Patent No. 6573086
 ; GENERAL INFORMATION:
 ; APPLICANT: EMALFARB, MARK AARON
 ; APPLICANT: BURLINGAME, RICHARD PAUL
 ; APPLICANT: OLSON, PHILIP TERRY
 ; APPLICANT: SINITSYN, ARKADY PANTELEIMONOVICK
 ; APPLICANT: PARRICHE, MARTINE
 ; APPLICANT: BOUSSON, JEAN CHRISTOPHE
 ; APPLICANT: PYNNONEN, CHRISTINE MARIE
 ; APPLICANT: PUNT, PETER JAN
 ; APPLICANT: VAN-ZEIJL, CORNELIA MARIA JOHANNA
 ; TITLE OF INVENTION: TRANSFORMATION SYSTEM IN THE FIELD OF FILAMENTOUS FUNGI
 ; FILE REFERENCE: 3123-4001
 ; CURRENT APPLICATION NUMBER: US/09/548,938A
 ; CURRENT FILING DATE: 2000-04-13
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 526
 ; TYPE: PRT
 ; ORGANISM: Chrysosporium lucknowense
 ; FEATURE:
 ; NAME/KEY: MOD_RES
 ; LOCATION: (249)
 ; OTHER INFORMATION: Variable amino acid
 ; FEATURE:
 ; NAME/KEY: MOD_RES
 ; LOCATION: (365)
 ; OTHER INFORMATION: Variable amino acid
 US-09-548-938A-10

Query Match 61.4%; Score 1681; DB 2; Length 526;
 Best Local Similarity 60.4%; Pred. No. 2.4e-121;
 Matches 311; Conservative 68; Mismatches 112; Indels 24; Gaps 10;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVIDANWRWTHATNSSTNCYDGNWTSSSTL 60
 |:|||| :| || ||| |:|:|:| ||: ||||| ||:|:|:|:| | ::
 Db 18 QNACTLTAENHPSLTWSKCTSGGSCTSVQGSITIDANWRWTHRTDSATNCYEGNKWDTSY 77

Qy 61 CPDNE[↓]XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 119
 || :|| ||:||| |:||||:|||||:|: |||: |:|:| ||| ||| ||
 Db 78 CSDGPSCASKCCIDGADYSSTYGITTSGNSLNLKFVTKGQYSTNIGSRTYLMESDTKYQM 137

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
 | |||||:||||| | ||||| |||||:| || | ||||| ||||| |||||
 Db 138 FQLLGNEFTFDVDVSNLGCGLNGALYFVSMDADGGMSKYSGNKAGAKYGTGYCDSQCPRD 197

Qy 180 LKFINGQ[↓]ANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
 |||||:|||| |: |:|:| | | :|||||:||||:|: ||||| :|| ||
 Db 198 LKFINGEANVENWQSS[↓]TNDANAGTGKYGSCCSEMDVWEANNMAAFTPHPCXVIGQSRCE 257

Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTQFETSGA 299
 || ||||| :|| | |||||:| || || :|| | | :|||:| ||||| : |
 Db 258 GDSCGGTYSTDRYAGICDPDGCDFNSYRQGNKTFYGGK--MTVDTTKKITVVTQFLKNSA 315

Qy 300 -----INRYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGGSSF-SDKGGLTQFK 353
 | :||| | : : || : | :| :| || : |||: |
 Db 316 GELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQDWCDRQKAAGDVTDXQDKGGMVQMG 375

Qy 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
 || :| |||||:| |: |||||:| : : || | ||:| | :|||:| |:|:|
 Db 376 KALAGPMVLVMSIWDHVNMLWLDSTWPI-DGAGKPGAERGACPTTSGVPAEVEAEAPN 434

Qy 414 AKVTFSTNIKFGPIGST--GNPSG--GNPPGGNPPGTTTTTRRP--ATTGSSPGPT----- 462
 : | |||:| ||||| | | | || | || :| :| | :||: | |||
 Db 435 SNVIFSNIRFGPIGSTVSGLPDGGSGNP--NPPVSSSTPVPSSSTTSSGSSGPTGGTGV 491

Qy 463 QSHYGQCGGIGYSGPTVCASGTTCCVLNPPYYSQCL 497
 || |||||:| || | | || | :||| |
 Db 492 AKHYEQCGGIGFTGPTQCESPYTCTKLNWYSQCL 526

RESULT 10

US-08-676-166A-3

; Sequence 3, Application US/08676166A

; Patent No. 5955270

; GENERAL INFORMATION:

; APPLICANT: Radford, Alan

; APPLICANT: Parish, John H.

; TITLE OF INVENTION: EXPLOITATION OF THE CELLULASE COMPLEX OF

; TITLE OF INVENTION: NEUROSPORA

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David A. Jackson, Esq.

; STREET: 411 Hackensack Ave, Continental Plaza, 4th

; STREET: Floor

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/676,166A

; FILING DATE: 15-JUL-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 1321-1-002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-487-5800

; TELEFAX: 201-343-1684

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 525 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: H. grisea

US-08-676-166A-3

; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-676-166A-2

Query Match 57.0%; Score 1561; DB 1; Length 516;
Best Local Similarity 57.5%; Pred. No. 4.2e-112;
Matches 294; Conservative 62; Mismatches 129; Indels 26; Gaps 10;

Qy	1	QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWTSSSTL	60
		: : : : :	
Db	18	QQAGTLTAKRHPSLTWQKCTRGGCPTLNT-TMVLNANWRWTHATSGSTKCYTGKNWQATL	76
		↓	
Qy	61	CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEF	120
		: : :	
Db	77	CPDGKSCAANCALDGADYTGTYGITGSGWSLTQLFVTD----NVGARAYLMADDTQYQML	132
Qy	121	TLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL	180
		: : : :	
Db	133	ELLNQELWFDVDMSNIPCGLNGALYLSAMDADGGMRKYPTNKAGAKYATGYCDAQCPRL	192
Qy	181	KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG	240
		: : : :	
Db	193	KYINGIANVEGWTPSTNDAN-GIGDHGSCCSEMDIWEANKVSTAFTPHPCTTIEQHMCEG	251
		↓	
Qy	241	DGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSGA-	299
		: : :	
Db	252	DSCGGTYSDDRYGVLCDADGCDFNSYRMGNTTFYGEKG--TVDTSSKFTVVTQFIKDSAG	309
Qy	300	----INRYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFKK	354
		: : : : : : : : :	
Db	310	DLAEIKAFYVQNGKVIENSQSNVDGVSGNSITQSFCKSQKTAFGDIDDFNKKGGLKQMGK	369
Qy	355	ATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNA	414
		: : : : : : :	
Db	370	ALAQAMVLVMSIWDDHAANMLWLDSTYP---VPKVPGAYRGS GP T TSGVPAEVDANAPNS	426
Qy	415	KVTFSNIKFGPI-----GSTGNPSGGNPPGGNPPGTTTTRRPATTTGSSP-GPTQSHY	466
		: : : : : : : : : :	
Db	427	KVAFSNIKFGHLGISPFSGSSGTPP-SNPSSASPTSSSTAKPSSSTASNPSTGAHAW	485
Qy	467	GQCGGIGYSGPTVCASGTTTCQVLNPPYYSQCL	497
		: : :	
Db	486	AQCGGIGFSGPTTCPEPYTCAKDHDIIYSQCV	516

RESULT 19
US-09-329-350-35
; Sequence 35, Application US/09329350
; Patent No. 6184019
; GENERAL INFORMATION:
; APPLICANT: Miettinen-Oinonen, Arja
; APPLICANT: Londesborough, John
; APPLICANT: Vehmaanper , Jari
; APPLICANT: Haakana, Heli

```

; APPLICANT: M ntyl , Arja
; APPLICANT: Lantto, Raija
; APPLICANT: Elovainio, Minna
; APPLICANT: Joutsjoki, Vesa
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko
; TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/329,350
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/841,636
; FILING DATE: 30-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,335
; FILING DATE: 17-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,926
; FILING DATE: 04-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,840
; FILING DATE: 28-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/732,181
; FILING DATE: 16-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI96/00550
; FILING DATE: 17-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Shea Jr., Timothy
; REGISTRATION NUMBER: 41,306
; REFERENCE/DOCKET NUMBER: 1716.0510006/MAC/TJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Melanocarpus albomyces
; STRAIN: ALKO4237
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..452
; OTHER INFORMATION: /label= 50K-cellulase-B
US-09-329-350-35

```



```

; APPLICATION NUMBER: US/08/841,636A
; FILING DATE: 30-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,335
; FILING DATE: 17-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,926
; FILING DATE: 04-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,840
; FILING DATE: 28-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/732,181
; FILING DATE: 16-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI96/00550
; FILING DATE: 17-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Timothy J. Shea, Jr.
; REGISTRATION NUMBER: 41,306
; REFERENCE/DOCKET NUMBER: 1716.0510005/MAC/TJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Melanocarpus albomyces
; STRAIN: ALKO4237
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..452
; OTHER INFORMATION: /label= 50K-cellulase-B
US-08-841-636A-35

```

```

Query Match          44.5%; Score 1219; DB 2; Length 452;
Best Local Similarity 51.9%; Pred. No. 8.2e-86;
Matches 221; Conservative 76; Mismatches 117; Indels 12; Gaps 10;

```

```

Qy      9  ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWSSSTLCPDNEXCA 68
      | |||||:|:| | | ||||| | | ||||| |:| | :|
Db     31  ENHPPLTWQRCTAPGNCQTVNAEVVIDANWRWLHDDNMQ-NCYDGNQWTNA-CSTATDCA 88

Qy     69  KNCCLDGAA-YASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQEFTLLGNE 126
      : | :| | | | ||| :|||:|:| |||: |||:| ||| || | | :| | |
Db     89  EKCMIEGAGDYLGTYGASTSGDALTLKFVTKHEYGTNVGSRFYLMNGPDKYQMFNLMGNE 148

Qy    127  FSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQ 186
      :|||:| : ||:| ||||:|: |||:| |:| |||:|||||:| | |||:| :
Db    149  LAFVDVLSTVECGINSALYFVAMEEDGGMASYPSNQAGARYGTGYCDAQCARDLKFVGGK 208

Qy    187  ANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGT 246
      ||:| | :|:| | | :| | :|||:|:|:|:| : | ||| ||| :| | |||
Db    209  ANIEGWKSSTSDPNAGVGPYGSCCAEIDVWESNAYAFAPHTPHACTTNEYHVCETTNCGGT 268

Qy    247  YSDNRYGGTCDPDGCDWNPYRLGNSTFYGPGSSFTLDTTKKLTVVVTQFETSGAINRYVQ 306
      ||:|:| | || :|||:||||:| ||| | ||||:| |||:| | :|:|:|
Db    269  YSEDRFAGKCDANGCDYNPYRMGNPDFYKGK--TLDTSRKFTTVSRFE-ENKLSQYFIQ 325

Qy    307  NG--VTFQQPNAELGSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFKKATSGGMVLV 363
      :| : | | :|:| :| : | : | | | |||

```


Qy	181	KFINGQANVEGWEPSSNNANTGIGGHGCCSEMDIWEANSISEALTPHPCTTVGQEICEG	240
		: : : : : :	
Db	198	-FINGLGNIE-----GKGACCNEMDIWEANARAQHIAPHPCSKAGPYLCEG	242
Qy	241	DGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSG--	298
		: : : : : :	
Db	243	AEC-----EFDGVCDKNGCAWNPYRVNVTDYEGEGAEFRVDTTRPFSVVTQFRAGGDA	295
Qy	299	-----AINRYVQNGVTFAQPNAEL-GSYSGNELNDDYCTAEAEFGSSFSKGGGLTQ	351
		: : : : : : : : :	
Db	296	GGGKLESIYRLFVQDGRVIESYVVDKPLPPTDRMTDEFCAAT----GAARFTELGAMEA	351
Qy	352	FKKATSGGMVLVMSLWDDYYANMLWLDDSTYPTNETSSSTPGAVRGSCSTSSGVPAQVESQS	411
		: : : :	
Db	352	MGDALTRGMVLALSIIWWSEGDNMNWLDS-----GEAGPCDPDEGNPSNIIRVQ	399
Qy	412	PNAKVTFSTNIKFGPIGSTGNPSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQSH	465
		: : : : : : : :	
Db	400	PDPEVVFSTNLRWGEIGST-YESAVDGPVGKKGKKGKAPAAA---GDGNGKEKSN	449

RESULT 33

US-08-709-974A-11

; Sequence 11, Application US/08709974A

; Patent No. 6117664

; GENERAL INFORMATION:

; APPLICANT: Sch lein, Martin

; APPLICANT: Rosholm, Peter

; APPLICANT: Nielsen, Jack Bech

; APPLICANT: Hansen, Svend Aage

; APPLICANT: von der Osten, Claus

; TITLE OF INVENTION: No. 6117664el Alkaline Cellulases

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,974A

; FILING DATE: 09-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valeta

; REGISTRATION NUMBER: 35,127y

; REFERENCE/DOCKET NUMBER: 4160.414-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 456 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-709-974A-11


```

; APPLICATION NUMBER: US/09/329,350
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/841,636
; FILING DATE: 30-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,335
; FILING DATE: 17-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,926
; FILING DATE: 04-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,840
; FILING DATE: 28-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/732,181
; FILING DATE: 16-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI96/00550
; FILING DATE: 17-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Shea Jr., Timothy
; REGISTRATION NUMBER: 41,306
; REFERENCE/DOCKET NUMBER: 1716.0510006/MAC/TJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 428 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Melanocarpus albomyces
; STRAIN: ALKO4237
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..428
; OTHER INFORMATION: /label= 50K-cellulase
US-09-329-350-33

```

```

Query Match          27.6%; Score 757; DB 2; Length 428;
Best Local Similarity 38.4%; Pred. No. 3.1e-50;
Matches 166; Conservative .59; Mismatches 151; Indels 56; Gaps 12;

```

```

Qy      9 ETHPPLTWQKCSSGGTCTQQTGSVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
      | | | | :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     28 EVHPQLTTFRCTKADGCQPRNTYIVLDSLHPVHQVDNDYNCGDWQKPNATACPDVESC 87

Qy     68 AKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLM-ASDTTYQEFTLLGNE 126
      |:| | :| :| | | | | :| :| :| :| :| :| :| :| :| :| :| :|
Db     88 ARNCIMEGVDPDYSQHGVTTSDTSLRLQLVDG--RLVTPRVYLLDETEHRYEMMHLTGQE 145

Qy    127 FSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRDLKFI 183
      |:| | :| | | | | | | | | | | | | | | | | | | | | | | | | |
Db    146 FTFEVDATKLPCGMNSALYLSEMDPTGARSE--LNPGGAYYGTGYCDAQCFVTP----FI 199

Qy    184 NGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGC 243
      || :| :| | | | | | | | | | | :| :| :| :| :| :| :| :|
Db    200 NGIGNIE-----GKGSCCNEMDIWEANSRATHVAPHTCNQTGLYMCEGAEC 245

Qy    244 GGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTVQFETSG----- 298
      | | | | | | | | :| :| :| :| :| :| :| :| :| :| :| :|
Db    246 -----EYDGVCDKDGCGWNPYRVNITDYGNSDAFRVDTRRPFTVVTVQFPADAEGRL 298

```

Qy	299	AINRYYYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATS	357
		: : : : : : :: : : : :	
Db	299	SIHRLYVQDGKVIESYVVDAPGLPRTDSLNDDEFCAAT----GAARYLDLGGTAGMGDAMT	354
		/	
Qy	358	GGMVLVMSLWDDYYANMLWLDDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVT	417
		: : :	
Db	355	RGMVLAMSIWWDSESGFMNWLDS-----GEAGPCLPDEGDPKNIVKVEPSPEVT	402
Qy	418	FSNIKFGPIGST	429
		: :::	
Db	403	YSNLRWGEIGST	414

RESULT 35

US-08-841-636A-33

; Sequence 33, Application US/08841636A

; Patent No. 6723549

; GENERAL INFORMATION:

; APPLICANT: Miettinen-Oinonen, Arja

; APPLICANT: Londesborough, John

; APPLICANT: Vehmaanper , Jari

; APPLICANT: Haakana, Heli

; APPLICANT: M ntyl , Arja

; APPLICANT: Lantto, Raija

; APPLICANT: Elovainio, Minna

; APPLICANT: Joutsjoki, Vesa

; APPLICANT: Paloheimo, Marja

; APPLICANT: Suominen, Pirkko

; TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND

; TITLE OF INVENTION: USES THEREOF

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Avenue, N.W., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/841,636A

; FILING DATE: 30-APR-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/005,335

; FILING DATE: 17-OCT-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/007,926

; FILING DATE: 04-DEC-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/020,840

; FILING DATE: 28-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/732,181

; FILING DATE: 16-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/FI96/00550

; FILING DATE: 17-OCT-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Timothy J. Shea, Jr.

; REGISTRATION NUMBER: 41,306

; REFERENCE/DOCKET NUMBER: 1716.0510005/MAC/TJS

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 428 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Melanocarpus albomyces
; STRAIN: ALKO4237
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..428
; OTHER INFORMATION: /label= 50K-cellulase
US-08-841-636A-33

```

```

Query Match          27.6%; Score 757; DB 2; Length 428;
Best Local Similarity 38.4%; Pred. No. 3.1e-50;
Matches 166; Conservative 59; Mismatches 151; Indels 56; Gaps 12;

```

```

Qy      9  ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
      ||| || :|:  | :| :|:|:  | :: || | | :| ||| |:|
Db     28  EVHPQLTTFRCCTKADGCQPRNTYIVLDSLHPVHQVDNDYNCGDWQKPNATACPDVESC 87

Qy     68  AKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLM-ASDTTYQEFTLLGNE 126
      |:| | ::|  | :||| | || : :  : | |:| |: :: | : | | |
Db     88  ARNCIMEGVDPDYSQHGVTTSDTSLRLQLQVLDG--RLVTPRVYLLDETEHRYEMMHLTGQE 145

Qy    127  FSEFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRDLKFI 183
      |:|:| | ::| |||:| |||  || | | :  | || ||||| |:| |  ||
Db    146  FTFEVDATKLP CGMNSALYLSEMDPTGARSE--LNPGGAYYGTGYCDAQCFVTP----FI 199

Qy    184  NGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTTPHPC'TTVGQEICEGDGC 243
      ||  |:|  | ||||:| ||||| | | : : || | | | :||| |
Db    200  NGIGNIE-----GKGSCCNEMDIWEANSRATHVAPHTCNQTGLYMCEGAEC 245

Qy    244  GGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGP GSSFTLDTTKKLTVVVTQFETSG----- 298
      | | || ||| |||||: | :|| :| :|| : |||||
Db    246  -----EYDGVCDKDGCGWNPYRVNITDYYGNSDAFRVDTRRPFTVVTQFPADAEGRL 298

Qy    299  AINRYVYVQNGVTFQQPNAEL-GSYSGNELND DYC'TAEAEFGGSSFS DKGGLTQFKKATS 357
      :|:| |||:|  :  : |  : |||:| |  | : : | ||  | :
Db    299  SIHRLYVQDGKVIESYVVDAPGLPRTDSL NDEFCAAT----GAARYLDLGGTAGMGDAMT 354

Qy    358  GGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVT 417
      |||| |||:| |  ||||  | |  | | :  | : :|
Db    355  RGMVLAMSIWWDESGFMWLDS-----GEAGPCLPDEGDPKNIVKVEPSPEVT 402

Qy    418  FSNIKFGPIGST 429
      :||::| ||||
Db    403  YSNLRWGEIGST 414

```

```

RESULT 36
US-08-709-974A-3
; Sequence 3, Application US/08709974A
; Patent No. 6117664
; GENERAL INFORMATION:
; APPLICANT: Sch lein, Martin
; APPLICANT: Rosholm, Peter
; APPLICANT: Nielsen, Jack Bech
; APPLICANT: Hansen, Svend Aage
; APPLICANT: von der Osten,Claus
; TITLE OF INVENTION: No. 6117664el Alkaline Cellulases

```

```

; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,974A
; FILING DATE: 09-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta
; REGISTRATION NUMBER: 35,127y
; REFERENCE/DOCKET NUMBER: 4160.414-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-709-974A-3

```

```

Query Match          27.4%; Score 750.5; DB 2; Length 409;
Best Local Similarity 38.3%; Pred. No. 9.3e-50;
Matches 171; Conservative 51; Mismatches 160; Indels 65; Gaps 15;

```

```

Qy      9 ETHPPLTWQKCSSGGTCTQQTGSVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
      | | | | : | : | | : | | | | | : | | | : |
Db      8 EQHPKLETYRCTKASGCKQTNYIVADAG---IHGIRRSAGCGDWGQKPNATACPDEASC 64

Qy     68 AKNCCLDGA---AYASTYGVTTSGNSLSIGFVTQSAQKN--VGARLYLMASD-TTYQEFT 121
      | | | | | | : | : | | | | | : | | | | : | :
Db     65 AKNCILSGMDSNAYKNA-GITTSGNKLRL----QQLINNQLVSPRVYLLEENKKKYEMLH 119

Qy    122 LLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PR 178
      | | | | | | : | : | | | | | | | | : | | | | | |
Db    120 LTGTEFSFDVEMEKLPCEMNGALYLSEMPQDGGKSTSRNSKAGAYYGAGYCDAQCYVTP- 178
      | | | | | | : | : | | | | | | | | : | | | | | |
Qy    179 DLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEIC 238
      | | | | | : | : | | | | | | : | : | | | : | |
Db    179 ---FINGVGNIK-----GQGVCCNELDIWEANSRATHIAPHPCSKPGLYGC 221
      | | | | | : | : | | | | | | : | : | | | : | |
Qy    239 EGDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTVQFETSG 298
      | | | | : | | | | | | : | | | | : | : | | | : |
Db    222 TGDECGSS-----GICDKAGCGWNHNRINVTDFYGRGKQYKVDSTRKFTVTSQFVANK 274

Qy    299 -----AINRYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEAEFGGSSFSKGGTLTQF 352
      : | : | : | : | : | | | | | : | : | | | |
Db    275 QGDIELHRHYIQDNKVIKSAVNISGPPKINFINDKYCAAT----GANEYMLRGGTKQM 330

Qy    353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
      | | | | | : | | | | | | | | : | | | : | |
Db    331 GDAMSRGMVLAMSVWWSEGDFMAWLDQ-----GVAGPCDATEGDPKNIVKVQP 378

Qy    413 NAKVTFSNIKFGPIGSTGNPSSGGNPPG 439
      | : | | | | : | | | | : | |

```


Qy	9	ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC	67
		: : : : :	
Db	8	EQHPKLETYRCTKASGCKKQTNIVADAG---IHGIRRSAGCGDWGQKPNATACPDEASC	64
Qy	68	AKNCCLDGA---AYASTYGVTTSGNSLSIGFVTQSAQKN--VGARLYLMASD-TTYQEFT	121
		: : : : : :	
Db	65	AKNCILSGMDSNAYKNA-GITTSNGKRLR----QQLINNQLVSPRVYLLEENKKKYEMLH	119
Qy	122	LLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PR	178
		: : : : :	
Db	120	LTGTEFSFDVEMEKLPCGMNGALYLSEMPQDGGKSTSRNSKAGAYYGAGYCDAQCYVTP-	178
Qy	179	DLKFINGQANVEGWEPSSNNANTGIGGGHSCCSEMDIWEANSISEALTTPHPCTTVGQEIC	238
		: : : : : : :	

```

Db      179 ---FINGVGNIK-----GQGVCCNELDIWEANSRATHIAPHPCSKPGLYGC 221
Qy      239 EGDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSG 298
      || || :      || || || || | : || || | : :| :| || :|| :
Db      222 TGDECGSS-----GICDKAGCGWNHNRINVTFDYGRGKQYKVDSTRKFTVTSQFVANK 274

Qy      299 -----AINRYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEAEFGGSSFSKGGTLTQF 352
      :| :| :| :      :      : |      :| || || |      | : :      || |
Db      275 QGD LIELHRHYIQDNKVIESAVVNISGPPKINFINDKYCAAT----GAN EYMRLGGTKQM 330
Qy      353 KKATSGGMVLVMSLWDDYYANMLWLWDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
      | | |||| || :|      | || |      | | | : | | :      |
Db      331 GDAMSRGMVLAMSVWWSEGDFMAWLDQ-----GVAGPCDATEGDPKNIVKVQP 378

Qy      413 NAKVTFSNIKFGPIGSTGNPSGGNPPG 439
      | : ||||| : | |||| :      ||
Db      379 NPEVTFSNIRIGEIGSTSSVKAPAYPG 405

```

RESULT 38

US-08-361-920-25

; Sequence 25, Application US/08361920

; Patent No. 5457046

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; TITLE OF INVENTION: or Hemicellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5457046o No. 5457046disk of No. 5457046th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/361,920

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/940,860

; FILING DATE: 28-OCT-1992

; APPLICATION NUMBER: DK 1158/90

; FILING DATE: 09-MAY-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/DK91/00124

; FILING DATE: 08-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 3435.204-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-867-0298

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 427 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-08-361-920-25

Query Match 27.4%; Score 750.5; DB 1; Length 427;
Best Local Similarity 38.3%; Pred. No. 9.9e-50;
Matches 171; Conservative 51; Mismatches 160; Indels 65; Gaps 15;

```
Qy      9  ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
      | | | | : | : | | : | | | | | | | : | | | | : |
Db     26  EQHPKLETYRCTKASGCKQTNYIVADAG---IHGIRRSAGCGDWGQKPNATACPDEASC 82

Qy     68  AKNCCLDGA---AYASTYGVTTSGNSLSIGFVTQSAQKN--VGARLYLMASD-TTYQEFT 121
      | | | | | | | | : | : | | | | | : | | | | | | : | : |
Db     83  AKNCILSGMDSNAYKNA-GITTSGNKRLR----QQLINNQLVSPRVYLLEENKKKYEMLQ 137

Qy    122  LLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PR 178
      | | | | | | | | : | : | | | | | | | | | | | | | | |
Db    138  LTGTEFSFDVEMEKLPCGMNGALYLSEMPQDGGKSTSRNSKAGAYYGAGYCDAQCYVTP- 196

Qy    179  DLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEIC 238
      | | | | | : | : | | | | | | | | | | | | : | | | |
Db    197  ---FINGVGNIK-----GQGVCCNELDIWEANSRATHIAPHPCSKPGLYGC 239

Qy    239  EGDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSG 298
      | | | | : | | | | | | | | : | | | | | : | : | | | : |
Db    240  TGDECGSS-----GFCDKAGCGWNHNRINVTDIFYGRGKQYKVDSTRKFTVTSQFVANK 292

Qy    299  -----AINRYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEAEFGGSSFSDKGGLTQF 352
      : | : | : | : | : | : | | | | | | | | : | : | | |
Db    293  QGD LIELHRHYIQDNKVIESAVVNISGPPKINFINDKYCAAT----GAN EYMLGGTKQM 348

Qy    353  KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
      | | | | | | | | : | | | | | | | | | | | | | : | |
Db    349  GDAMSRGMVLAMSVWWSEGDFMAWLDQ-----GVAGPCDATEGDPKNIVKVQP 396

Qy    413  NAKVTFSNIKFGPIGSTGNPSGGNPPG 439
      | : | | | | | : | | | | : | |
Db    397  NPEVTFSNIRIGEIGSTSSVKAPAYPG 423
```

RESULT 39

US-08-479-939-25

; Sequence 25, Application US/08479939

; Patent No. 5686593

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; TITLE OF INVENTION: or Hemicellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5686593o No. 5686593disk of No. 5686593th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/479,939

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE: 22-DEC-1994
; APPLICATION NUMBER: US 07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 427 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-479-939-25

Query Match 27.4%; Score 750.5; DB 1; Length 427;
Best Local Similarity 38.3%; Pred. No. 9.9e-50;
Matches 171; Conservative 51; Mismatches 160; Indels 65; Gaps 15;

Qy	9	ETHPPLTWQKCSSGGTCTQQTGSSVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC	67
		: : : : :	
Db	26	EQHPKLETYRCTKASGCKKQNTYIVADAG---IHGIRRSAGCGDWGQKPNATACPDASC	82
Qy	68	AKNCCLDGA---AYASTYGVTTSGNSLSIGFVTQSAQKN--VGARLYLMASD-TTYQEFT	121
		: : : : :	
Db	83	AKNCILSGMDSNAYKNA-GITTSNKLRL---QQLINNQLVSPRVYLLEENKKKYEMLO	137
Qy	122	LLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PR	178
		: : : :	
Db	138	LTGTEFSFDVEMEKLPCEMNGALYLSEMPQDGGKSTSRNSKAGAYYGAGYCDAQCYVTP-	196
Qy	179	DLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEIC	238
		: : : : : :	
Db	197	---FINGVGNK-----GQGVCCNELDIWEANSRATHIAPHPCSKPGLYGC	239
Qy	239	EGDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFQFETSG	298
		: : : : :	
Db	240	TGDECGSS-----GFCDKAGCGWNHNRINVTDFYGRGKQYKVDSTRKFTVTSQFVANK	292
Qy	299	-----AINRYYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEAEFGGSSFSKGGTLTQF	352
		: : : : : : : : : : : :	
Db	293	QGDLELHRHYIQDNKVIESAVVNISSPPKINFINDKYCAAT---GANEYMRLGGTKQM	348
Qy	353	KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP	412
		: : : : :	
Db	349	GDAMSRGMVLAMSVWWSEGDFMAWLDQ-----GVAGPCDATEGDPKNIVKVQP	396
Qy	413	NAKVTFSTNIFGPIGSTGNPSGGNPPG	439
		: : :	
Db	397	NPEVTFSTNIRIGEIGSTSSVKAPAYPG	423

RESULT 40
US-08-483-432-25
; Sequence 25, Application US/08483432
; Patent No. 5763254

```

; GENERAL INFORMATION:
; APPLICANT: Woeldike, Helle F.
; APPLICANT: Hagen, Frederick
; APPLICANT: Hjort, Carsten M.
; APPLICANT: Sven, Hastrup
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
; TITLE OF INVENTION: or Hemicellulose
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5763254o No. 5763254disk of No. 5763254th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,432
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE:
; APPLICATION NUMBER: US 07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 427 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-432-25

```

```

Query Match          27.4%; Score 750.5; DB 1; Length 427;
Best Local Similarity 38.3%; Pred. No. 9.9e-50;
Matches 171; Conservative 51; Mismatches 160; Indels 65; Gaps 15;

```

```

Qy      9 ETHPPLTWQKCSSGGTCTQQTGSGVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
      | | | | : | : | : | | | | | | | | : | | | : |
Db     26 EQHPKLETYRCTKASGCKKQNTNYIVADAG---IHGIRRSAGCGDWGQKPNATACPDEASC 82

Qy     68 AKNCCLDGA---AYASTYGVTTSGNSLSIGFVTQSAQKN--VGARLYLMASD-TTYQEFT 121
      | | | | | | | | : | : | | | | | : | | | | : | : | : |
Db     83 AKNCILSGMDSNAYKNA-GITTSGNKLRL----QQLINNQLVSPRVYLLEENKKKYEMLQ 137

Qy    122 LLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPNTNAGAKYGTGYCDSQC---PR 178
      | | | | | | | : : | | | : | | | | | | | : | | | | | | : |
Db    138 LTGTEFSFDVEMEKLP CGMNGALYLSEMPQDGGKSTSRNSKAGAYYGAGYCDACQCYVTP- 196

Qy    179 DLKFINGQANVEGWEPSSNNANTGIGGGHSCCSEMDIWEANSISEALTPHPCTTVGQEIC 238
      | | | | | : : | | | : | | | | | : : | | | : | | |

```

```

Db      197 ---FINGVGNIK-----GQGVCCNELDIWEANSRATHIAPHPCSKPGLYGC 239
Qy      239 EGDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSG 298
      || || :      || || || || | : || || | : :||:| || :|| :
Db      240 TGDECGSS-----GFCDKAGCGWNHNRINVTDIFYGRGKQYKVDSTRKFTVTSQFVANK 292

Qy      299 -----AINRYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEAEFGGSSFSKGGGLTQF 352
      :||:|:|: : : : | :|| || | : : || |
Db      293 QGD LIELHRHYIQDNKVIESAVVNISGPPKINFINDKYCAAT----GAN EYMRLGGTKQM 348

Qy      353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
      | | |||| ||:| | || | | | : | | : |
Db      349 GDAMSRGMVLAMSVWWSEGDFMAWLDQ-----GVAGPCDATEGDPKNIVKVQP 396

Qy      413 NAKVTFSNIKFGPIGSTGNPSGGNPPG 439
      | :|||||: | |||| :      ||
Db      397 NPEVTFSNIRIGEIGSTSSVKAPAYPG 423

```

RESULT 41

US-08-709-974A-6

; Sequence 6, Application US/08709974A

; Patent No. 6117664

; GENERAL INFORMATION:

; APPLICANT: Sch lein, Martin

; APPLICANT: Rosholm, Peter

; APPLICANT: Nielsen, Jack Bech

; APPLICANT: Hansen, Svend Aage

; APPLICANT: von der Osten,Claus

; TITLE OF INVENTION: No. 6117664el Alkaline Cellulases

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,974A

; FILING DATE: 09-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valeta

; REGISTRATION NUMBER: 35,127y

; REFERENCE/DOCKET NUMBER: 4160.414-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 411 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-709-974A-6

Query Match 27.1%; Score 741.5; DB 2; Length 411;

Best Local Similarity 38.0%; Pred. No. 4.6e-49;

Matches 170; Conservative 52; Mismatches 162; Indels 63; Gaps 15;

Qy	9	ETHPPLTWQKCSSGGTCTQQTGSVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC	67
Db	8	EQHPKLETYRCTKASGCKKQTNVIADAGIHGIRQKNGA-GCGDWGQKPNATACPDEASC	66
Qy	68	AKNCCLDGA---AYASTYGVTTSGNSLSIGFVTQSAQKN--VGARLYLMASD-TTYQEFT	121
Db	67	AKNCILSGMDSNAYKNA-GITTSGNKLRL----QQLINNQLVSPRVYLLEENKKKYEMLH	121
Qy	122	LLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PR	178
Db	122	LTGTEFSFDVEMEKLPCGMNGALYLSEMPQDGGKSTSRNSKAGAYYGAGYCDAQCYVTP-	180
Qy	179	DLKFINGQANVEGWEPSSNNANTGIGGGHSGCCSEMDIWEANSISEALTTPHPCTTVGQEIC	238
Db	181	---FINGVGNIK-----GQGVCCNELDIWEANSRATHIAPHPCSKPGLYGC	223
Qy	239	EGDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSG	298
Db	224	TGDECGSS-----GICDKAGCGWNHNRINVTDFYGRGKQYKVDSTRKFTVTSQFVANK	276
Qy	299	-----AINRYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEAEFGGSSFSDKGGLTQF	352
Db	277	QGDIELHRHYIQDNKVIESAVVNIISGPPKINFINDKYCAAT----GANEYMRLGGTKQM	332
Qy	353	KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSSTPGAVRGSCSTSSGVPAQVESQSP	412
Db	333	GDAMSRGMVLAMSVWWSEGDFMAWLDQ-----GVAGPCDATEGDPKNIVKVQP	380
Qy	413	NAKVTFSTNIKFGPIGSTGNPSSGGNPPG	439
Db	381	NPEVTFSNIRIGEIGSTSSVKAPAYPG	407

RESULT 42
 US-09-069-632-1
 ; Sequence 1, Application US/09069632
 ; Patent No. 6261828
 ; GENERAL INFORMATION:
 ; APPLICANT: Lund, Henrik
 ; TITLE OF INVENTION: A Process For Combined Desizing
 ; TITLE OF INVENTION: And Stone-Washing of Dyed Denim
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 6261828o No. 6261828disk of No. 6261828th America, Inc.
 ; STREET: 405 Lexington Avenue
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10174
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/069,632
 ; FILING DATE: 29-APR-1998
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/DK96/00469
 ; FILING DATE: 15-NOV-1996
 ; APPLICATION NUMBER: 1278/95
 ; FILING DATE: 15-NOV-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gregg, Valeta
 ; REGISTRATION NUMBER: 35,127
 ; REFERENCE/DOCKET NUMBER: 4588.204-US

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-069-632-1

```

Query Match 27.0%; Score 739.5; DB 2; Length 415;
Best Local Similarity 37.6%; Pred. No. 6.7e-49;
Matches 170; Conservative 58; Mismatches 155; Indels 69; Gaps 16;

Qy	9	ETHPLTWQKCSSGGTCTQQTGSGVVIDANRWTHATN--SSTNCYD-GNTWSSTLCPDNE	65
		: : : : :	
Db	8	EVHPQLTTFRCTKRGGCKPATNFIVLDSLSPHIAHRAEGLGPGGCGDWGNPPPKDVC PDVE	67
		↓	
Qy	66	XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTT--YQEFTL	122
		: : : : : : : : :	
Db	68	SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL	124
Qy	123	LGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRD	179
		: : : : :	
Db	125	TGFETFTDVDATKLPCGMNSALYLSEMHPTGAKSKY--NPGGAYYGTGYCDAQCFVTP--	180
Qy	180	LKFINGQANVEGWEPSSNNANTGIGGGHGCCSEMDIWEANSISEALTPHPCTTVGQEICE	239
		: : : : :	
Db	181	--FINGLGNIE-----GKGSCCNEMDIWEANSRASHVAPHTCNKKGLYLCE	224
Qy	240	GDGCGGTYS DNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS--	297
		: : : : : : : :	
Db	225	GEECA-----FEGVCDKNGCGWNNYRVNVTDDYGRGEEFKVNTLKPFTVVVTQFLANRR	277
Qy	298	---GAINRYYVQNGVTFQQ--PNAELGSYSGNELNDYCTAEAEFGGSSFSFDKGGLTQF	352
		: : : : : : : : :	
Db	278	GKLEKIHRFYVQDGKVIESFYTNKEGV PYT-NMIDDEFCEAT----GSRKYMELGATQGM	332
Qy	353	KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP	412
		: : : : : :	
Db	333	GEALTRGMVLAMSIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQVEP	380
Qy	413	NAKVTFFSNIKFGPIGST-----GNPSGGNPP	438
		: : : : :	
Db	381	FPEVTYTNLRWGEIGSTYOEVOVKPKPKPGHGP	412

RESULT 43

US-08-709-974A-5

; Sequence 5, Application US/08709974A

; Patent No. 6117664

; GENERAL INFORMATION:

; APPLICANT: Sch lein, Martin

; APPLICANT: Rosholm, Peter

; APPLICANT: Nielsen, Jack Bech

APPLICANT: Hansen, Svend Aage

; APPLICANT: von der Osten, Claus

TITLE OF INVENTION: No. 6117664e1 Alkaline Cellulases

```
; NUMBER OF SEQUENCES: 11
```

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,974A
; FILING DATE: 09-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta
; REGISTRATION NUMBER: 35,127y
; REFERENCE/DOCKET NUMBER: 4160.414-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-709-974A-5

Query Match 26.9%; Score 737.5; DB 2; Length 402;
Best Local Similarity 38.2%; Pred. No. 9.2e-49;
Matches 167; Conservative 57; Mismatches 150; Indels 63; Gaps 15;

Qy	9	ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE	65
		: : : : : : : :	
Db	8	EVHPQLTTFRCTKRGCKPATNFIVLDSLHPIHRAEGLGPGGCGDWGNPPPKDVC PDVE	67
Qy	66	XCAKNCCLDGAAYASTYGVTTSNLSIGFVTQSAQKNVGARLYLMASDTT---YQEFTL	122
		: : : : : : : : :	
Db	68	SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL	124
Qy	123	LGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRD	179
		: : : : :	
Db	125	TGFEFTFDVDATKLPCGMNSALYLSEMHPTGAKSKY--NPGGAYYGTGYCDAQCFVTP--	180
Qy	180	LKFINGQANVEGWEPSSNNANTGIGGGHGCCSEMDIWEANSISEALTPHPCTTVGQEICE	239
		: : : : : :	
Db	181	--FINGLGNIE-----GKGSCCNEMDIWEANSRASHVAPHTCNKKGLYLCE	224
Qy	240	GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS--	297
		: : : : : : : : :	
Db	225	GEECA-----FEGVCDKNGCGWNNYRVNVTDYGRGEEFKVNTLKPFTVVVTQFLANRR	277
Qy	298	---GAINRYYVQNGVTFQQ--PNAELGSYSGNELNDYCTAEEAEFGGSSFSDKGGLTQF	352
		: : : : : : : : : :	
Db	278	GKLEKIHRYFVQDGKVIESFYTNKEGVPT-NMIDDEFCEAT----GSRKYMELGATQGM	332
Qy	353	KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSSTPGAVRGSCSTSSGVPAQVESQSP	412
		: : : : : : :	
Db	333	GEALTRGMVLAMSIIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQVEP	380
Qy	413	NAKVTFSTNIKFGPIGST	429
		: : : : :	
Db	381	FPEVTTYTNLRWGEIGST	397

```

; GENERAL INFORMATION:
; APPLICANT: Woeldike, Helle F.
; APPLICANT: Hagen, Frederick
; APPLICANT: Hjort, Carsten M.
; APPLICANT: Sven, Hastrup
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
; TITLE OF INVENTION: or Hemicellulose
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5457046o No. 5457046disk of No. 5457046th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-361-920-27

```

```

Query Match          26.9%; Score 737.5; DB 1; Length 435;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 169; Conservative 59; Mismatches 155; Indels 69; Gaps 16;

```

```

QY      9  ETHPPLTWQKCSSGGTCTQQTGSGVVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
      | ||| :|: | | :|:|: | | | | :|||
Db     28  EVHPQLTTFRCTKRGGCKPATNFIVLDSLHPIHRAEGLPGGCGDWGNPPPKDVC PDVE 87

QY     66  XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASD TT---YQEFTL 122
      :| ||| :| | ||||:| || : : : | :|: | | :|
Db     88  SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL 144

QY    123  LGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
      | ||:| ||| :| ||||:| || | | | | | | :| | |
Db    145  TGFEFTFDVDATKLP CGMNSALYLSEMHPTGAKSKY--NSGGAYYGTGYCDAQCFVTP-- 200

QY    180  LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      |||| :| | ||||:| |||| | | : : | | | :|
Db    201  --FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKKGLYLCE 244

```

Qy	240	GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS--	297
Db	245	GEECA-----FEGVCDKNGCGWNNYRVNVTDDYYGRGEEFKVNTLKPFTVVVTQFLANRR	297
Qy	298	---GAINRYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEAEFGSSFSDDKGGTLTQF	352
Db	298	GKLEKIHRFYVDGKVIESFYTNKEGVPT-NMIDDEFCEAT----GSRKYMELGATQGM	352
Qy	353	KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP	412
Db	353	GEALTRGMVLAMSIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQVEP	400
Qy	413	NAKVTFSTNIKFGPIGST-----GNPSGGNPP	438
Db	401	FPEVTTYTNLRWGEIGSTYQEVQKPKPKPGHGP	432

RESULT 45

US-08-479-939-27

; Sequence 27, Application US/08479939

; Patent No. 5686593

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; TITLE OF INVENTION: or Hemicellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5686593o No. 5686593disk of No. 5686593th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/479,939

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/361,920

; FILING DATE: 22-DEC-1994

; APPLICATION NUMBER: US 07/940,860

; FILING DATE: 28-OCT-1992

; APPLICATION NUMBER: DK 1158/90

; FILING DATE: 09-MAY-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/DK91/00124

; FILING DATE: 08-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 3435.204-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-867-0298

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 435 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-08-479-939-27

Query Match 26.9%; Score 737.5; DB 1; Length 435;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 169; Conservative 59; Mismatches 155; Indels 69; Gaps 16;

```
Qy      9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
      | | | | : | : | | : | : | | | | | : | | | |
Db     28 EVHPQLTTFRCKTRGGCKPATNFIVLDSLSPHIAEGLGPGCGDWGNPPPKDVC PDVE 87

Qy     66 XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTT--YQEFTL 122
      : | | | | : | : | | | | : | | : : | | | : | |
Db     88 SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL 144

Qy    123 LGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
      | | | : | | | : | | | : | | | | | | | | : | | |
Db    145 TGFEFTFDVDATKLPCGMNSALYLSEMHPTGAKSKY--NSGGAYYGTGYCDAQCFVTP-- 200

Qy    180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      | | | | : | : | | | | : | | | | | | | | : | |
Db    201 --FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKKGLYLCE 244

Qy    240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTQFETS-- 297
      | : | : | : | | : | | | | : | | | | : | | | | :
Db    245 GEECA-----FEGVCDKNGCGWNNYRVNVTDYGRGEEFKVNTLKPFTVVTQFLANRR 297

Qy    298 ---GAINRYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEAEFGGSSFSDKGGLTQF 352
      | : | : | | : | : | | : | : | : | | : : |
Db    298 GKLEKIHRFYVDGKVIESFYTNKEGVPT-NMIDDEFCEAT----GSRKYMELGATQGM 352

Qy    353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
      : | : | | | | : | | | | | | | | | | | : : |
Db    353 GEALTRGMVLAMSIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQVEP 400

Qy    413 NAKVTFSNIKFGPIGST-----GNPSGGNPP 438
      : | | : | : | | | | | | | : | |
Db    401 FPEVTTYTNLRWGEIGSTYQEVQKPKPKPGHGP 432
```

RESULT 46

US-08-483-432-27

; Sequence 27, Application US/08483432

; Patent No. 5763254

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; TITLE OF INVENTION: or Hemicellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5763254o No. 5763254disk of No. 5763254th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/483,432

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE:
; APPLICATION NUMBER: US 07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-432-27

Query Match 26.9%; Score 737.5; DB 1; Length 435;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 169; Conservative 59; Mismatches 155; Indels 69; Gaps 16;

Qy	9	ETHPPLTWQKCSSGGTCTQQTGSSVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE	65
		: : : : :	
Db	28	EVHPQLTTFRC TKRGCKPATNFIVLDSLHPHRAEGLGPGGCGDWGNPPPKDVC PDVE	87
Qy	66	XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTT---YQEFTL	122
		: : : : : : :	
Db	88	SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL	144
Qy	123	LGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRD	179
		: : :	
Db	145	TGFEFTFDVDATKLP CGMNSALYLS E MHPTGAKSKY--NSGGAYYGTGYCDAQCFVTP--	200
Qy	180	LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE	239
		: : :	
Db	201	--FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKKGLYLCE	244
Qy	240	GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS--	297
		: : : : :	
Db	245	GEECA-----FEGVCDKNGCGWNNYRVNVTDYGRGEEFKVNTLKPFTVVVTQFLANRR	297
Qy	298	---GAINRYYVQNGVTFQQ--PNAELGSYSGNELNDYCTAEAEFGGSSFSKGGLTQF	352
		: : : : : : : :	
Db	298	GKLEKIHRFYVQDGKVIESFYTNKEGV PYT-NMIDDEFCEAT----GSRKYMELGATQGM	352
Qy	353	KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP	412
		: : : : :	
Db	353	GEALTRGMVLAMSIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQVEP	400
Qy	413	NAKVTFSTNIKFGPIGST-----GNPSGGNPP	438
		: : : : : :	
Db	401	FPEVTTYTNLRWGEIGSTYQEVQKPKPKPGHGP	432

```

; GENERAL INFORMATION:
; APPLICANT: Lund, Henrik
; TITLE OF INVENTION: A Process For Combined Desizing
; TITLE OF INVENTION: And Stone-Washing of Dyed Denim
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6261828o No. 6261828disk of No. 6261828th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/069,632
; FILING DATE: 29-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK96/00469
; FILING DATE: 15-NOV-1996
; APPLICATION NUMBER: 1278/95
; FILING DATE: 15-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 4588.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-069-632-3

```

```

Query Match          26.9%; Score 737.5; DB 2; Length 435;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 169; Conservative 59; Mismatches 155; Indels 69; Gaps 16;

```

```

QY      9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
      | | | | : | : | | : | : | | | | : | | |
Db     28 EVHPQLTTFRCTKRGGCKPATNFIVLDSLHPIHRAEGLPGGCGDWGNPPPKDVC PDVE 87
      /
QY     66 XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTT---YQEFTL 122
      : | | | | : : | | | | | : | | : : | | : | | |
Db     88 SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL 144

QY    123 LGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
      | | : | | | : : | | | : | | | | | | | : | | | | | |
Db    145 TGFEFTFDVDATKLP CGMNSALY LSEMHPTGAKSKY--NSGGAYYGTGYCDAQCFVTP-- 200
      |      |
QY    180 LKFINGQANVEGWEPSSNNANTGIGGHSGCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      | | | | : | | | | | | | | | | : : | | | : | |
Db    201 --FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKKGLYLCE 244
      \ /
QY    240 GDGCGGTYS DNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLT VVTQFETS-- 297
      | : | : | | : | | | | | : | : | | | | | :
Db    245 GEECA-----FEGVCDKNGCGWNNYRVNVTDYGRGEEFKVNTLKPFTVVTQFLANRR 297

QY    298 ---GAINRYYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEAEFGGSSFSDKGGLTQF 352

```

```

      |:|:| | | | : | | | : | : | : | | : : |
Db      298 GKLEKIHRFYVQDGKVIESFYTNKEGVPYT-NMIDDEFCEAT----GSRKYMELGATQGM 352
      /
QY      353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPQAVESQSP 412
      : | : | | | | | : | | | | | | : | | : : |
Db      353 GEALTRGMVLAMSIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQVEP 400

QY      413 NAKVTFSNIKFGPIGST-----GNPSGGNPP 438
      : | | : | : : | | | | | | : | : |
Db      401 FPEVTTYTNLRWGEIGSTYQEVQKPKPKPGHGP 432

```

RESULT 48

US-08-709-974A-1

```

; Sequence 1, Application US/08709974A
; Patent No. 6117664
; GENERAL INFORMATION:
;   APPLICANT: Sch lein, Martin
;   APPLICANT: Rosholm, Peter
;   APPLICANT: Nielsen, Jack Bech
;   APPLICANT: Hansen, Svend Aage
;   APPLICANT: von der Osten,Claus
;   TITLE OF INVENTION: No. 6117664el Alkaline Cellulases
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.
;     STREET: 405 Lexington Avenue, 64th Floor
;     CITY: New York
;     STATE: New York
;     COUNTRY: United States of America
;     ZIP: 10174-6401
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/709,974A
;     FILING DATE: 09-SEP-1996
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Gregg, Valeta
;     REGISTRATION NUMBER: 35,127y
;     REFERENCE/DOCKET NUMBER: 4160.414-US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 212-867-0123
;     TELEFAX: 212-878-9655
;   INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 402 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein

```

US-08-709-974A-1

```

Query Match          26.8%; Score 735.5; DB 2; Length 402;
Best Local Similarity 38.0%; Pred. No. 1.3e-48;
Matches 166; Conservative 58; Mismatches 150; Indels 63; Gaps 15;

```

```

QY      9  ETHPPLTWQKCSSGGTCTQQTGSVVIDANNRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
      | | | | : | : | | : | : | | | | : | | |
Db      8  EVHPQLTTFRCTKRGGCKPATNFIVLDSLHPIHRAEGLPGGCGDWGNPPPKDVC PDVE 67

QY      66 XCAKNCCLDGAAYASTYGVTTSGNLSIGFVTQSAQKNVGARLYLMASDTT---YQEFTL 122
      : | | | | : : | | | | | : | | : : | | : |
Db      68 SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL 124

```

Qy 123 LGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
 | | | : | | | : : | | | : | | | | | : | | | | | : | | | |
 Db 125 TGFEFTFDVDATKLPCGMNSALYLSEMHPTGAKSKY--NSGGAYYGTGYCDAQCFVTP-- 180

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
 | | | | : | | | | : | | | | : | | | : | | | : | |
 Db 181 --FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKKGLYLCE 224

Qy 240 GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS-- 297
 | : | : | | : | | : | | | | : | | | : | | | | :
 Db 225 GEECA-----FEGVCDKNGCGWNNYRVNVTDYGRGEEFKVNTLKPFTVVVTQFLANRR 277

Qy 298 ---GAINRYYVQNGVTFFQQ--PNAELGSYSGNELNDDYCTAEAEFGGSSFSKGGTLTQF 352
 | : | : | | : | : | | : | : | : | | : : |
 Db 278 GKLEKIHRFYVDGKVIESFYTNKEGVPT-NMIDDEFCEAT----GSRKYMELGATQGM 332

Qy 353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
 : | : | | | | : | | | | | | | | : | : | : |
 Db 333 GEALTRGMVLAMSIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQVEP 380

Qy 413 NAKVTFSNIKFGPIGST 429
 : | | : | : : | | | |
 Db 381 FPEVTTYTNLRWGEIGST 397

RESULT 49

US-08-709-974A-4

; Sequence 4, Application US/08709974A

; Patent No. 6117664

; GENERAL INFORMATION:

; APPLICANT: Sch lein, Martin

; APPLICANT: Rosholm, Peter

; APPLICANT: Nielsen, Jack Bech

; APPLICANT: Hansen, Svend Aage

; APPLICANT: von der Osten,Claus

; TITLE OF INVENTION: No. 6117664el Alkaline Cellulases

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,974A

; FILING DATE: 09-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valeta

; REGISTRATION NUMBER: 35,127y

; REFERENCE/DOCKET NUMBER: 4160.414-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 415 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-08-709-974A-4

Query Match 26.8%; Score 735.5; DB 2; Length 415;
Best Local Similarity 37.4%; Pred. No. 1.4e-48;
Matches 169; Conservative 58; Mismatches 156; Indels 69; Gaps 16;

```
Qy      9  ETHPPLTWQKCSSGGTCTQQTGSVVIDANNRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
      | | | | : | : | | : | : | | | | : | | |
Db      8  EVHPQLTTFRCTKRGCKPATNFIVLDSLHPIHRAEGLGPGGCGDWGNPPPKDVC PDVE 67

Qy     66  XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTT---YQEFTL 122
      : | | | | : : | | | | | : | | : : | : | : | | |
Db     68  SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL 124

Qy    123  LGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
      | | | : | | | : : | | | | | | | | | | | | | | | |
Db    125  TGFEFTFDVDATKLPCGMNSALYLSEMHPTGAKSKY--NPGGAYYGTGYCDAQCFVTP-- 180

Qy    180  LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      | | | | : | : | | | | : | | | | | | | | : : | |
Db    181  --FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKKGLYLCE 224

Qy    240  GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTQFETS-- 297
      | : | : | : | | : | | | | | : | : | | | : | | | :
Db    225  GEECA-----FEGVCDKNGCGWNNYRVNVTDYGRGEEFKVNTLKPFTVVTQFLANRR 277

Qy    298  ---GAINRYVQNGVTFQQ--PNAELGSYSGNELNDYCTAEAEFGGSSFSDKGGLTQF 352
      | : | : | | : : | | : | : | : | | : : |
Db    278  GKLEKIHRFYVQDGKVIESFYTNKEGVPT-NMIDDEFCEAT----GSRKYMELGATQGM 332

Qy    353  KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
      : | : | | | | : | | | | | | | | | : | : | : |
Db    333  GEALTRGMVLAMSIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQVEP 380

Qy    413  NAKVTFSNIKFGPIGST-----GNPSGGNPP 438
      : | | : | : : | | | | | | | : | : |
Db    381  FPEVYTNLRWGEIGSTYQEVQKPKPKPGHGP 412
```

RESULT 50

US-08-709-979A-3

; Sequence 3, Application US/08709979A

; Patent No. 5912157

; GENERAL INFORMATION:

; APPLICANT: Claus von der Osten

; APPLICANT: Martin Sch lein

; TITLE OF INVENTION: No. 5912157el Alkaline Cellulases

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5912157o No. 5912157disk of No. 5912157th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,979A

; FILING DATE: 09-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4160.404-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-709-979A-3

Query Match 26.5%; Score 725.5; DB 1; Length 402;
Best Local Similarity 37.7%; Pred. No. 7.7e-48;
Matches 166; Conservative 59; Mismatches 146; Indels 69; Gaps 16;

Qy	9	ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSST-----NCYD-GNTWSSTLCP	62
		: : : : : : : : :	
Db	8	EVHPQLTTFRCKTRGGCKPATNFIV---DLSLSHPIHRAEGLPGGCGDWGNPPPKDVCP	64
Qy	63	DNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTT---YQE	119
		: : : : : : : : : : :	
Db	65	DVESCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEM	121
Qy	120	FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---	176
		: :	
Db	122	LHLTGFEFTFDVDATKLPCGMNSALYLSEMHPTGAKSKY--NPGGAYYGTGYCDAQCFVT	179
Qy	177	PRDLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQE	236
		: : : :	
Db	180	P-----FINGLGNIE-----GKGSCCNEMDIWEANSRASHVAPHTCNKKGLY	221
Qy	237	ICEGDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFET	296
		: : : : : : :	
Db	222	LCEGEECA-----FEGVCDKNGCGWNNYRVNVTDYGRGEEFKVNTLKPFTTVVTQFLA	274
Qy	297	S-----GAINRYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEAEFGGSSFSKGGGL	349
		: : : : : : : :	
Db	275	NRRGKLEKIHRFYVQDGKVIESFYTNKEGVPYT-NMIDDEFCEAT----GSRKYMELGAT	329
Qy	350	TQFKKATSGGMVLVMSLWDDYYANMLWLDDSTYPTNETSSTPGAVRGSCSTSSGVPQAVES	409
		: : : : :	
Db	330	QGMGEALTRGMVLAMSIIWWDQGGNMEWLDH-----GEAGPCAKGEGAPSNIVQ	377
Qy	410	QSPNAKVTFSTNIKFGPIGST	429
		: : : :	
Db	378	VEPFPEVITYTNLRWGEIGST	397

RESULT 51
US-08-833-642A-5
; Sequence 5, Application US/08833642A
; Patent No. 5883066
; GENERAL INFORMATION:
; APPLICANT: Ivan M. A. J. Herbots et al.
; TITLE OF INVENTION: Liquid Detergent Compositions
; TITLE OF INVENTION: Containing Cellulase and Amine
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jackie Ann Zurcher
; ADDRESSEE: Dinsmore & Shohl LLP
; STREET: 255 E. Fifth Street
; STREET: 1900 Chemed Center
; CITY: Cincinnati
; STATE: Ohio

```

; COUNTRY: USA
; ZIP: 45202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,642A
; FILING DATE: April 8, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Zurcher, J. A.
; REGISTRATION NUMBER: P42,251
; REFERENCE/DOCKET NUMBER: CM551C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 977-8377
; TELEFAX: (513) 977-8141
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-833-642A-5

```

```

Query Match          25.7%; Score 705.5; DB 1; Length 415;
Best Local Similarity 37.3%; Pred. No. 2.8e-46;
Matches 163; Conservative 59; Mismatches 152; Indels 63; Gaps 15;

```

```

Qy      9 ETHPPLTWQKCSSGGTCTQQTGSVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
      | | | | : | : | | : | : | | | | : | | |
Db      8 EVHPQLTTFRCTKRGGCKPATNFIVLDSLHP IHRAEGLGPGGCGDHGNPPPKDVC PDVE 67

Qy     66 XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASD TT---YQEFTL 122
      : | | | | : | : | | | | : | | : : | | | : |
Db     68 SCAKNCIMEGIPDYSQYGVTTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL 124

Qy    123 LGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
      | | | : | | | : | | | : | | | | | | | : | |
Db    125 TGFEFTFDVDATKLPCGMNSALYLFENHPTGAKSKY--NSGGAYYGTGYCDAQCFVTP-- 180

Qy    180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      | | | | : | : | | | | : | | | | | | : | |
Db    181 --FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKKGLYLCE 224

Qy    240 GDGCGGTYS DNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLT VVVTQFETS-- 297
      | : | : | | : | | : | | : | | : | | | | :
Db    225 GEECA-----FEGVCDKNGCGYNNYRVNVTDYGRGEEFKVNTLKPFTVVVTQFLANRR 277

Qy    298 ---GAINRYYVQNGVTFQQ--PNAELGSYSGNELND DYTAEAEFGGSSFS DKGGLTQF 352
      | : | : | | : | : | | : | : | | : | : |
Db    278 GRLEKIH RFYVQDGKVI ESFYTNKEGV PYT-NMIDDEFCEAT----GSRKYMELGATQGM 332

Qy    353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
      : | : | | | | : | | | | | | | : | : |
Db    333 GEALTRGMVLAMSIWWDQGGNMENLDH-----GEAGPCAKGEGAPSNIVQVEP 380

Qy    413 NAKVTFSNIKFGPIGST 429
      : | | : | : : | | | |
Db    381 FPEVTTYTNLRWGEIGST 397

```

```

RESULT 4
S38794
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (Humicola grisea)
N;Alternate names: beta-glucan cellobiohydrolase; exoglucanase
C;Species: Humicola grisea var. thermoidea

```

C;Datè: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S38794; S08240; A45869
R;Radford, A.
submitted to the EMBL Data Library, June 1991
A;Reference number: S38794
A;Accession: S38794
A;Molecule type: DNA
A;Residues: 1-525 <RAD>
A;Cross-references: UNIPROT:P15828; UNIPARC:UPI000012BE0F; EMBL:X17258; NID:g2760;
PIDN:CAA35159.1; PID:g2761
A;Note: this is a revision to the sequence from reference S08240
R;de Oliveira Azevedo, M.; Radford, A.
Nucleic Acids Res. 18, 668, 1990
A;Title: Sequence of cbh-1 gene of Humicola grisea var. thermoidea.
A;Reference number: S08240; MUID:90175006; PMID:2308855
A;Accession: S08240
A;Molecule type: DNA
A;Residues: 1-299, 'H', 301-525 <DEO>
A;Cross-references: UNIPARC:UPI00001729F6; EMBL:X17258
A;Note: the authors translated the codon CAG for residue 87 as His
A;Note: this sequence has been revised in reference S38794
R;Azevedo, M.; de, O.; Felipe, M.S.S.; Astolfi-Filho, S.; Radford, A.
J. Gen. Microbiol. 136, 2569-2576, 1990
A;Title: Cloning, sequencing and homologies of the cbh-1 (exoglucanase) gene of Humicola grisea
var. thermoidea.
A;Reference number: A45869; MUID:91178527; PMID:2127803
A;Accession: A45869
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-20, 'R', 22-34, 'K', 36-86, 'H', 88-141, 'V', 143-157, 'Y', 159-237, 'QQH', 241-244, 'I', 246-
299, 'H', 301-525 <AZE>
A;Cross-references: UNIPARC:UPI00001729F7; GB:M64588; GB:X17258
A;Note: this sequence has been revised. See entry S08240
C;Genetics:
A;Gene: cbh-1
A;Introns: 138/1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;494-525/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 60.3%; Score 1652; DB 1; Length 525;
Best Local Similarity 57.3%; Pred. No. 1.5e-91;
Matches 294; Conservative 76; Mismatches 121; Indels 22; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWTWSSTL 60
| ||:| :| || |:| ||::|| | :| :||| | : |||| | | : ::
Db 19 QQACSLTTERHPSLSWNKCTAGGQCQTVQASITLDSNWRWTHQVSGSTNCYTGKWDTSI 78

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE 119
| | ::||:||||:| | ||||:|:|:|:|: ||: |||:| || | : ||
Db 79 CTDAKSCAQNCCVDGADYTSTYGITTINGDSLKLFVTKGQHSTNVGSRTYLMGDGEDKYQT 138

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYP'TNTAGAKYGTGYCDSQCPRD 179
| |||||:| |||| | : ||||| ||||| ||||:|:| | ||||| ||||| : |||||
Db 139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMDADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
: |||||:|:| || | :|: | | | :|:| ||||| ||||:|: | ||||| :| | ||
Db 199 IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE 258

Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS-- 297
|| |||||:| || | |||||:| || || :|| | | :|||:| |||||
Db 259 GDSCGGTYSNERYAGVCDPDGCDFNSYRQGNKTFYKGK--MTVDTTKKITVVVTQFLKDAN 316

Qy 298 ---GAINRYVQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFK 353
| | |:|:|:| : : || : |:| : : || | : |||:|
Db 317 GDLGEIKRFYVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

```

Qy      354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
      || :| |||||:||||: :|||||||:| : : ||| ||:| |:|||||:|:|:|
Db      377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy      414 AKVTFSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQS 464
      : | ||||:||||| | :|||| | |||: |||| :| ||
Db      436 SNVVFVNIRFGPIGSTVAGLPGAGNGGNGGNPP--PPTTTTSSAPATTTTASAGPKAG 492

Qy      465 HYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
      : |||||:| || | || :||||
Db      493 RWQQCGGIGFTGPTQCEEPYICTKLNDWYSQCL 525

```

RESULT 6

S42093

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - *Neurospora crassa*

C;Species: *Neurospora crassa*

C;Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C;Accession: S42093

R;Taleb, F.; Radford, A.

submitted to the EMBL Data Library, February 1994

A;Description: Cloning sequencing and homologies of the CBH-1 (exocellobiohydrolase) gene of *Neurospora crassa*.

A;Reference number: S42093

A;Accession: S42093

A;Molecule type: DNA

A;Residues: 1-516 <TAL>

A;Cross-references: UNIPROT:P38676; UNIPARC:UPI000011D714; EMBL:X77778; NID:g456657;

PIDN:CAA54815.1; PID:g456658

C;Genetics:

A;Introns: 227/3

C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homology

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

F;485-516/Domain: fungal cellulose-binding domain homology <FCB>

```

Query Match          57.0%; Score 1561; DB 2; Length 516;
Best Local Similarity 57.5%; Pred. No. 4.1e-86;
Matches 294; Conservative 62; Mismatches 129; Indels 26; Gaps 10;

```

```

Qy      1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
      | | || :| |||||: || | | :|:|||||||: || || | | :|
Db      18 QQAGTLTAKRHPSLTWQKCTRGGCPTLNT-TMVLNANWRWTHATSGSTKCYTGKNWQATL 76

Qy      61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEF 120
      ||| :| || |||| | |||:| || ||:| || | |||| |||| | |
Db      77 CPD GKSCAANCALDGADYTGTYGITGSGWSLTQLFVTD----NVGARAYLMADDTQYQML 132

Qy      121 TLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 180
      || | ||||:| :||||||| :|||||: |||| |||| | |||:| ||||
Db      133 ELLNQELWFDVDSNIPCGLNGALYLSAMDADGGMRYPTNKAGAKYATGYCDAQCPRL 192

Qy      181 KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG 240
      |:||| ||||| ||:|:| || | ||||| |||| :| | |||||:| :|||
Db      193 KYINGIANVEGWTPSTNDAN-GIGDHGSCCSEMDIWEANKVSTAFTPHPCTTIEQHMCEG 251

Qy      241 DCGGTYSDNRYGGTCDPDGCDWNPYRLGNSTSFYGPSSFTLDTTKKLTVVVTQFETSGA- 299
      | |||||:| || | ||||:| ||:|:| || | |:|:| | ||||| |
Db      252 DSCGGTYSDDRYGVLCADGCDFN SYRMGNNTTFYEGEK--TVDTSSKFTVVVTQFIKDSAG 309

Qy      300 ----INRYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFKK 354
      | :||| | : : : ||| : :| :|: || | |: |||| | |
Db      310 DLAEIKAFYVQNGKVIENSQSNVDGVSNSITQSFKKSQKTAFGDIDDFNKKGGLKQMGK 369

Qy      355 ATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNA 414
      | : |||||:| |: ||||| |||| | ||| ||| |:|||||:|:|:|
Db      370 ALAQAMVLVMSIWDDHAANMLWLDSTYP---VPKVP GAYRSGSPTTSGVPAEVDANAPNS 426

```

Qy 415 KVTFSNIKFGPI-----GSTGNPSGGNPPGGNPPGTTTTTRRPATTTGSSP-GPTQSHY 466
 || ||||| : ||:| || |::| : |:| |:| :|:
 Db 427 KVAFSNIKFGHLGISPFSGGSSGTPP-SNPSSSASPTSSTAKPSSTSTASNPSGTGAHAW 485

Qy 467 GQCGGIGYSGPTVCASGTTTCQVLNPYYSQL 497
 |||||:|||| | || : ||||:
 Db 486 AQCGGIGFSGPTTCPEPYTCAKDHDIIYSQCV 516

RESULT 13

JE0313

exoglucanase (EC 3.2.-.-) - imperfect fungus (*Humicola grisea*)

C;Species: *Humicola grisea*

C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 09-Jul-2004

C;Accession: JE0313

R;Takashima, S.; Iikura, H.; Nakamura, A.; Hidaka, M.; Masaki, H.; Uozumi, T.

J. Biochem. 124, 717-725, 1998

A;Title: Isolation of the gene and characterization of the enzymatic properties of a major exoglucanase of *Humicola grisea* without a cellulose-binding domain.

A;Reference number: JE0313; MUID:98429588; PMID:9756616

A;Accession: JE0313

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-451 <TAK>

A;Cross-references: UNIPROT:O93780; UNIPARC:UPI000005E865; DDBJ:AB003105

C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homology

C;Keywords: glycosidase; hydrolase

Query Match 45.4%; Score 1243.5; DB 2; Length 451;
 Best Local Similarity 52.0%; Pred. No. 3.3e-67;
 Matches 226; Conservative 85; Mismatches 113; Indels 11; Gaps 9;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSSVIDANWRWTHATNSSTNCYDGNWTSSSTL 60
 | | : : | | : ||::| | | | | | | | | | : | | : | | : |
 Db 23 QQAGTITAENHPRMTWKRCSPGNCQTVQGEVVIDANWRWLH--NNGQNCYEGNKWTSQ- 79

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQE 119
 | : ||: | ||| | ||| : |||: ||:: ||: | : | | | : | |
 Db 80 CSSATDCAQRCALDGANYQSTYGASTSGDSLTLKFVTKHEYGTNIGSRFYLMANQNKYQM 139

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
 |||: |||: |||: ||: ||: | ||||: ||: |||: ||: | ||||| |||: || |
 Db 140 FTLMNNEFAFDVDLSKVECGINSALYFVAMEEDGGMASYPSNRAGAKYGTGYCDAQCARD 199

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCSEMDIWEANSISEALTPHPCTTVGQ-EIC 238
 |||| |: ||: || | ||: | | : | | : ||: ||: ||: : | || | : : | |
 Db 200 LKFIGGKANIEGWRPSTNDPNAGVGPMGACCAEIDVWESNAYAYAFTPHACGSKNRYHIC 259

Qy 239 EGDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFQFETSG 298
 | : |||||: ||: | | : |||: |||: || | | | : | | | : | :
 Db 260 ETNNCGGTYSDDRFAGYCDANGCDYNPYRMGNKDFYKGK--TVDTNRKFTVVSFRFERN- 316

Qy 299 AINRYYVQNGVTFFQQPNAEL-GSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFKKAT 356
 : :: ||: | : | : : : | | : | : | : | : | : | : | : | :
 Db 317 RLSQFFVQDGRKIEVPPPTWGPLNSADITPELCDAQFRVFDDRNRFAETGGFDALNEAL 376

Qy 357 SGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKV 416
 : |||||: |||: |||: |||: || | : | | | | | : ||||: ||: | : | : |
 Db 377 TIPMVLVMSIWDDHHSNMLWLDSSYPP-EKAGLPGGDRGPCPTTSGVPAEVEAQYPDAQV 435

Qy 417 TFSNIKFGPIGSTGN 431
 : |||: ||||| |
 Db 436 VWSNIRFGPIGSTVN 450


```

Db      139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198
QY      180 LKFINGQANVEGWEPSSNNANTGIGGGHSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
      :|||||:|:| | | :|: | | | :|:|||||||:|:| | | | | :| | |
Db      199 IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCIIIGQSRCE 258
QY      240 GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS-- 297
      || |||||: | | | |||||:| | | | :|| | | | :|||:| | | |
Db      259 GDSCGGTYSNERYAGVCDPDGCDFNSYRQGNKTFYKKG--MTVD'TTKKITVVVTQFLKDN 316
QY      298 ---GAINRYYYQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFK 353
      | | | |:| | | : | : | | | : | : | | | : | | | : | | |
Db      317 GDLGEIKRFYVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376
QY      354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
      || :| |||||:| | | : |||||:| | | : | : || | | | | :| | | | | :| |
Db      377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435
QY      414 AKVTFSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQS 464
      : | ||||:| | | | | | | | : |||| | | |||: | | | | : | | |
Db      436 SNVVFNSNIRFGPIGSTVAGLPAGNGGNGGNPP--PPTTTTSSAPATTTTASAGPKAG 492
QY      465 HYGQCGGIGYSGPTVCASGTTTCQVLNPYYSQCL 497
      : |||||:| | | | | | | | || | | :||| |
Db      493 RWQCGGIGFTGPTQCEEPYTCTKLNDWYSQCL 525

```

RESULT 15

```

GUX1_HUMGT
ID      GUX1_HUMGT      STANDARD;      PRT;      525 AA.
AC      P15828;
DT      01-APR-1990, integrated into UniProtKB/Swiss-Prot.
DT      01-FEB-1996, sequence version 2.
DT      07-FEB-2006, entry version 55.
DE      Exoglucanase 1 precursor (EC 3.2.1.91) (Exoglucanase I)
DE      (Exocellobiohydrolase I) (1,4-beta-cellobiohydrolase) (Beta-
DE      glucancellobiohydrolase).
GN      Name=CBH-1;
OS      Humicola grisea var. thermoidea.
OC      Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX      NCBI_TaxID=5528;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX      MEDLINE=90175006; PubMed=2308855;
RA      de Oliveira Alzevedo M., Radford A.;
RT      "Sequence of cbh-1 gene of Humicola grisea var. thermoidea.";
RL      Nucleic Acids Res. 18:668-668(1990).
CC      -!- FUNCTION: The biological conversion of cellulose to glucose
CC      generally requires three types of hydrolytic enzymes: (1)
CC      Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC      Exocellobiohydrolases that cut the dissaccharide cellobiose from
CC      the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
CC      glucosidases which hydrolyze the cellobiose and other short cello-
CC      oligosaccharides to glucose.
CC      -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC      in cellulose and cellotetraose, releasing cellobiose from the non-
CC      reducing ends of the chains.
CC      -!- SIMILARITY: Belongs to the glycosyl hydrolase 7 (cellulase C)
CC      family.
CC      -!- SIMILARITY: Contains 1 CBM1 (fungal-type carbohydrate-binding)
CC      domain.
CC      -----
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      -----
DR      EMBL; X17258; CAA35159.1; -; Genomic_DNA.

```


Title: - US-10-804-785-2 (Thr66 deleted)

RESULT 170

ABJ26900

ID ABJ26900 standard; protein; 460 AA.

XX

AC ABJ26900; .

XX

DT 23-OCT-2003 (revised)

DT 08-MAY-2003 (first entry)

XX

DE Cellobiohydrolase I activity protein SEQ ID No 52.

XX

KW Cellobiohydrolase; enzyme; DNA shuffling; ethanol; biomass;

KW cellobiohydrolase I; EC 3.2.1.91.

XX

OS Coprinopsis cinerea.

XX

PN WO2003000941-A2.

XX

PD 03-JAN-2003.

XX

PF 26-JUN-2002; 2002WO-DK000429.

XX

PR 26-JUN-2001; 2001DK-00001000.

XX

PA (NOVO) NOVOZYMES AS.

XX

PI Lange L, Wu W, Aubert D, Landvik S, Schnorr KM, Clausen IG;

XX

DR WPI; 2003-278244/27.

DR N-PSDB; ABT23538.

XX

PT New polypeptide with cellobiohydrolase I activity, useful in producing

PT ethanol from biomass.

XX

PS Claim 4; Page 175-177; 199pp; English.

XX

CC The invention relates to a novel polypeptide comprising: part of any of
CC 21 amino acid sequences; an amino acid sequence at least 70% identical to
CC a polypeptide encoded by a cellobiohydrolase gene; an amino acid sequence
CC at least 80% identical to the polypeptide encoded by 21 nucleotide
CC sequences; a polypeptide encoded by a nucleotide sequence which
CC hybridises with a probe selected from complementary strands of 55
CC nucleotide sequences; or a fragment of the aforementioned structures. The
CC polynucleotides of the invention are useful in a method of DNA shuffling.
CC The polypeptides are useful in a method for producing ethanol from
CC biomass comprising contacting the biomass with the polypeptides. This
CC sequence represents a protein with cellobiohydrolase I activity of the
CC invention. (Updated on 23-OCT-2003 to standardise OS field)

XX

SQ Sequence 460 AA;

Query Match 46.0%; Score 1259; DB 6; Length 460;

Best Local Similarity 53.5%; Pred. No. 1.9e-73;

Matches 234; Conservative 64; Mismatches 117; Indels 22; Gaps 9;

Qy 8 SETHPPLTWQKCSSGGTC-TQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPD-NE 65

Db 26 AENHPRLPWQRCTRNGGCQTVSNGQVVLNANWRWLHVTGTYTNCYTGNSTVCSNPTT 85

Qy 66 CAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNE 125

Db 86 CAQRCALEGANYQQTYGITNGDALTIKFLTRSQQTNVVGARVYLMENENRYQMFNLLNKE 145

Qy 126 FSHFDVDSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQ 185

Db	146	FTFDVDVSKVPCGINGALYFIQMDADGGMSKQPNNRAGAKYGTGYCDSQCPRDIKFIDGV	205
Qy	186	ANVEGWEPSSNNANTGIGGGHSCCSEMDIWEANSISEALTPHPCTTV--GQEICEGDGC	242
		: : : :	
Db	206	ANSADWTPSETDPNAGRGRYGICCAEMDIWEANSISNAYTPHPCRTQNDGGYQRCGRDC	265
Qy	243	GGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFET-----SG	297
		: : : : : : : :	
Db	266	----NQPRYEGLCDPDGCDYNPFRMGNKDFYGP GK--TVDTNRKMTVVVTQFITHDNTDTG	319
Qy	298	A---INRYYVQNGVTFQQPNAELGSY--SGNELNDDYCTAEAEAFGG-SSFSDKGGLTQF	351
		: : : : : : : :	
Db	320	TLVDIRRLYVQDGRVIANPPTNFPGLMPAHD SITEQFCTDQKNLFGDYSSFARDGGLAHM	379
Qy	352	KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP	411
		: : : : : : : : : : : :	
Db	380	GRSLAKGHVLALSIWNHDGAHMLWLDSNYPTDADPNKPGIARGTCPTTGGTPRETEQNHP	439
Qy	412	NAKVTFSNIKFGPIGST	428
		: :	
Db	440	DAOVIFSNIKFGDIGST	456

RESULT 171

ID AAR94351 standard; protein: 451 AA.

AC AAR94351:

DT 29-AUG-1996 (first entry)

DE Humicola insolens cellulase.

KW Cellulase; detergents; textile auxiliaries; feed additives;
KW digestive agents; host cell; recombinant production.

OS Humicola insolens.

FH	Key	Location/Qualifiers
----	-----	---------------------

FT /label

```
FT /label= mat peptide
```

PN JP08056663-A.

PD 05-MAR-1996.

PF 29-AUG-1994; 94JP-00203564.

PR 29-AUG-1994; 94JP-00203564.

PA (MEIJ) MEIJI SEIKA KAISHA LTD.

DR WPI; 1996-182296/19.

PT Humicola insolens cellulase - used as main component in detergents,
PT textile auxiliaries, feed additives and digestive agents.

PS Claim 2; Page 9-10; 16pp; Japanese.

CC The present sequence is *H. insolens* cellulase, which is used as the main
CC component in detergents, textile auxiliaries, feed additives and
CC digestive agents. A host cell transformed with a vector contg. the
CC cellulase DNA, can be used for the recombinant prodn. of the cellulase

SQ Sequence 451 AA;

Query Match 45.6%; Score 1250; DB 2; Length 451;
Best Local Similarity 52.1%; Pred. No. 7e-73;
Matches 226; Conservative 84; Mismatches 114; Indels 10; Gaps 8;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
 | | : : | | : | : | | | | | | | | : | | : | | : |
Db 23 QQAGTITAENHPRMTWKRCSPGNCTVQGEVVIDANWRWLH--NNGQNCYEGNKWTSQC 80

Qy 61 CPDNECAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQEF 119
 : | : | | | | | | | : | : | : | : | | : | | : | | |
Db 81 SSATDCAQRCAIDGANYQSTYGASTSGDSLTLKFVTKHEYGTNIGSRFYLMANQNKYQMF 140

Qy 120 TLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 179
 | | : | | : | | | : | : | | | : | : | | | : | : | | | : | | |
Db 141 TLMNNEFAFDVDSLKVECGINSALYFVAMEEDGGMASYPSNRAGAKYGTGYCDAQCARDL 200

Qy 180 KFIGQANVEGWEPSSNNANTGIGGHGCCSEMDIWEANSISEALTPHPCTTVGQ-EICE 238
 | | | | : | : | | | | | : | : | | | : | : | : | : | : | : | : | | |
Db 201 KFIGGKANIEGWRPSTNDPNAGVGPAGACCAEIDVWESNAYAYAFTPHACGSKNRYHICE 260

Qy 239 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETSGA 298
 : | | | | | : | : | | : | : | : | : | | | | | : | : | : | : | :
Db 261 TNNCGGTYSDDRFAGYCDANGCDYNPYRMGNKDFYKGK--TVDTNRKFTVVSFRFERN-R 317

Qy 299 INRYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFKKATS 356
 : : : : | : | : | : | : : : : | : | : | : | : | : | : | :
Db 318 LSQFFVQDGRKIEVPPPTWPGLPNSADITPELCDAQFRVFDDRNRFAETGGFDALNEALT 377

Qy 357 GGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPQVESQSPNAKVT 416
 | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 378 IPMVLVMSIWDHHSNMLWLDSSYPP-EKAGLPGGDRGPCPTTSGVPAEVEAQYPNAQVV 436

Qy 417 FSNIKFGPIGSTGN 430
 : | | : | | | | | | |
Db 437 WSNIRFGPIGSTVN 450

RESULT 172

AAW44852

ID AAW44852 standard; protein; 451 AA.

XX

AC AAW44852;

XX

DT 31-JUL-1998 (first entry)

XX

DE Humicola insolens cellulase NCE1 protein.

XX

KW Humicola insolens; NCE1; NCE2; NCE4; cellulase; expression vector;
KW promoter; signal sequence; terminator; amylase; lipase; protease;
KW phytase.

XX

OS Humicola insolens.

XX

FH Key Location/Qualifiers

FT Peptide 1. .21

FT /label= signal

FT Protein 22. .451

FT /label= Cellulase_NCE1

XX

PN WO9803667-A1.

XX

PD 29-JAN-1998.

XX

PF 24-JUL-1997; 97WO-JP002560.

XX

PR 24-JUL-1996; 96JP-00195070.
XX
PA (MEIJ) MEIJI SEIKA KAISHA LTD.
XX
PI Moriya T, Murashima K, Aoyagi K, Sumida N, Watanabe M, Hamaya T;
PI Koga J, Kono T, Murakami T;
XX
DR WPI; 1998-120786/11.
DR N-PSDB; AAV19376.
XX
PT Mass production of proteins and peptides in Humicola species - using
PT expression vector containing the promoter, signal sequence and/or
PT terminator from the Humicola insolens NCE1 or NCE2 gene.
XX
PS Claim 8; Page 34-39; 63pp; Japanese.
XX
CC The present sequence represents the Humicola insolens cellulase NCE1
CC protein from the present invention. The present invention describes a
CC method for the mass production of proteins and peptides in Humicola
CC species, especially in Humicola insolens, using an expression vector
CC which comprises the promoter, signal sequence and/or terminator
CC regulatory sequences from the NCE1 or NCE2 gene of H. insolens. These are
CC available in the plasmids pM3-1 (Escherichia coli JM109/pM3-1, FERM BP-
CC 5971) (for NCE1) and pM14-1 (E. coli JM109/pM14-1, FERM BP-5972) (for
CC NCE2). The vector also contains a marker gene such as an antibiotic
CC resistance gene (e.g. the destomycin resistance gene from Streptomyces
CC rimofaciens). Proteins which can be expressed using this system include
CC cellulase, amylase, lipase, protease, phytase and other enzymes. Specific
CC expression vectors of the invention are pMKD01 (for Humicola NCE3
CC cellulase gene), pEGD01 (for Humicola NCE4 cellulase gene) and pIED02
CC (for Humicola NCE4 cellulase gene). The expression system allows the
CC efficient production of proteins and peptides in a Humicola host. Using
CC the expression system high amounts of protein (>4.5 g/l) can be obtained
XX
SQ Sequence 451 AA;

Query Match 45.6%; Score 1250; DB 2; Length 451;
Best Local Similarity 52.1%; Pred. No. 7e-73;
Matches 226; Conservative 84; Mismatches 114; Indels 10; Gaps 8;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSSVIDANNRWTHATNSSTNCYDGNWSSSTL 60
| | : : | | : : | | | | | | | | : | : | : | : |
Db 23 QQAGTITAENHPRMTWKRCSPGNCQTVQGEVVIDANNRWLH--NNGQNCYEGNKWTSQC 80
Qy 61 CPDNECAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQEF 119
: | : | | | | | | | : | : | : | : | : | : | : |
Db 81 SSATDCAQRCAIDGANYQSTYGASTSGDSLTLKFVTKHEYGTNIGSRFYLMANQNKYQMF 140
Qy 120 TLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 179
| | : | : | : | : | : | : | | | | : | : | : | : | : | : | : |
Db 141 TLMNNEFAFDVDSLKVECGINSALYFVAMEEDGGMASYPNIRAGAKYGTGYCDAQCARDL 200
Qy 180 KFINGQANVEGWEPSSNNANTGIGGHGCCSEMDIWEANSISEALTTPHCTTVGQ-EICE 238
| | | : | : | : | | : | : | : | : | : | : | : | : | : | : |
Db 201 KFIGGKANIEGWRPSTNDPNAGVGPAGACCAEIDVWESNAYAYAFTPHACGSKNRYHICE 260
Qy 239 GDGCGGTYSNDRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTQFETSGA 298
: | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 261 TNNCGGTYSDDRFAGYCDANGCDYNPYRMGNKDFYKGK--TVDTNRKFTVVSFRFERN-R 317
Qy 299 INRYVQNGVTFQQPNAEL-GSYSGNELNDYCTAEAEFGG-SSFSDKGGTLQFKKATS 356
: : : : | : | : | : : : | : | : | : | : | : | : | : |
Db 318 LSQFFVQDGRKIEVPPPTWPGPLNSADITPELCDAQFRVFDNRNFAETGGFDALNEALT 377
Qy 357 GGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVT 416
| | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 378 IPMVLVMSIWDHHSNMLWLDSSYPPEKAGLPGGDRGPCPTTSGVPAEVEAQYPNAQVV 436

QY 417 FSNIKFGPIGSTGN 430
:||||:||||||| |
Db 437 WSNIRFGPIGSTVN 450

RESULT 173

ABJ26887

ID ABJ26887 standard; protein; 451 AA.

XX

AC ABJ26887;

XX

DT 08-MAY-2003 (first entry)

XX

DE Cellobiohydrolase I activity protein SEQ ID No 6.

XX

KW Cellobiohydrolase; enzyme; DNA shuffling; ethanol; biomass;

KW cellobiohydrolase I; EC 3.2.1.91.

XX

OS Scytalidium sp.

XX

PN WO2003000941-A2.

XX

PD 03-JAN-2003.

XX

PF 26-JUN-2002; 2002WO-DK000429.

XX

PR 26-JUN-2001; 2001DK-00001000.

XX

PA (NOVO) NOVOZYMES AS.

XX

PI Lange L, Wu W, Aubert D, Landvik S, Schnorr KM, Clausen IG;

XX

DR WPI; 2003-278244/27.

DR N-PSDB; ABT23505.

XX

PT New polypeptide with cellobiohydrolase I activity, useful in producing

PT ethanol from biomass.

XX

PS Claim 4; Page 119-121; 199pp; English.

XX

CC The invention relates to a novel polypeptide comprising: part of any of
CC 21 amino acid sequences; an amino acid sequence at least 70% identical to
CC a polypeptide encoded by a cellobiohydrolase gene; an amino acid sequence
CC at least 80% identical to the polypeptide encoded by 21 nucleotide
CC sequences; a polypeptide encoded by a nucleotide sequence which
CC hybridises with a probe selected from complementary strands of 55
CC nucleotide sequences; or a fragment of the aforementioned structures. The
CC polynucleotides of the invention are useful in a method of DNA shuffling.
CC The polypeptides are useful in a method for producing ethanol from
CC biomass comprising contacting the biomass with the polypeptides. This
CC sequence represents a protein with cellobiohydrolase I activity of the
CC invention

XX

SQ Sequence 451 AA;

Query Match 45.5%; Score 1245; DB 6; Length 451;
Best Local Similarity 51.8%; Pred. No. 1.5e-72;
Matches 225; Conservative 85; Mismatches 114; Indels 10; Gaps 8;

QY 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNWTSSSTL 60

Db 23 QQAGTITAENHPRMTWKRCSPGNCQTVQGEVVIDANWRWLH--NNGQNCYEGNKWTSQC 80

QY 61 CPDNECAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQEF 119

Db 81 SSATDCAQRCAIDGANYQSTYGASTSGDSLTLKFVTKHEYGTNIGSRFYLMANQNKYQMF 140

Qy	120	TLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL	179
		: : : :	
Db	141	TLMNNEFAFDVDSLKVECGINSALYFVAMEEDGGMASYPSNRAGAKYGTGYCDAQCARDL	200
Qy	180	KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQ-EICE	238
		: : : : : : :	
Db	201	KFIGGKANIEGWRPSTNDPNAGVGPMPGACCAEIDVWESNAYAYAFTPHACGSKNRYHICE	260
Qy	239	GDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFQFETSGA	298
		: : : : : : : :	
Db	261	TNNCGGTYSDDRFAGYCDANGCDYNPYRMGNKDFYKGKG--TVDTNRKFTVVSFRFERN-R	317
Qy	299	INRYYVQNGVTFQQPNAEL-GSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFKKATS	356
		: : : : : : : : : : : : : :	
Db	318	LSQFFVQDGRKIEVPPPTWPGLPNSADITPELCDAQFRVFDDRNRFETGGFDALNEALT	377
Qy	357	GGMVLVMSLWDDYYANMLWLWDSTYPTNETSSTPGAVRGSCSTSSGVPQAVESQSPNAKVT	416
		: : : : : : :	
Db	378	IPMVLVMSIWDDHHSNMLWLWDSSYPP-EKAGLPGGDRGPCPTTSGVPAEVEAQYPDAQVV	436
Qy	417	FSNIKFGPIGSTGN	430
		: :	
Db	437	WSNIRFGPIGSTVN	450

RESULT 9
 US-09-463-712C-10
 ; Sequence 10, Application US/09463712C
 ; Patent No. 6558937
 ; GENERAL INFORMATION:
 ; APPLICANT: DSM, N.V.
 ; APPLICANT: Gielkens, Marcus
 ; APPLICANT: Vesser, Jacob
 ; APPLICANT: De Graaff, Leendert
 ; TITLE OF INVENTION: CELLULOSE DEGRADING ENZYMES OF
 ; TITLE OF INVENTION: ASPERGILLUS
 ; FILE REFERENCE: 24615-20135.00
 ; CURRENT APPLICATION NUMBER: US/09/463,712C
 ; CURRENT FILING DATE: 2000-04-04
 ; PRIOR APPLICATION NUMBER: PCT/EP98/05047
 ; PRIOR FILING DATE: 1998-07-31
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 10
 ; LENGTH: 536
 ; TYPE: PRT
 ; ORGANISM: Aspergillus niger
 US-09-463-712C-10

Query Match 61.1%; Score 1673.5; DB 2; Length 536;
 Best Local Similarity 59.6%; Pred. No. 2.3e-122;
 Matches 308; Conservative 61; Mismatches 125; Indels 23; Gaps 7;

Qy	1	QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL	60
		: : : : :	
Db	22	QQVGTYTTETHPSLTWQTCTSDGSCCTNDGEVVIDANWRWVHSTSSATNCYTGNEWDTSI	81
Qy	61	CPDN-ECAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEF	119
		: : : : : :	
Db	82	CTDDVTCAANCALDGATYEATYGVTTSGSELRLNFVTQGSCKNIGSRLYLMSDDSNYELF	141
Qy	120	TLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL	179
		: :	
Db	142	KLLGQEFTFDVDVSNLPCGLNGALYFVAMDADGGTSEYSGNKAGAKYGTGYCDSQCPRDL	201
Qy	180	KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG	239

Db 202 KFINGEANC DGWEPSSNNVNTGVGDHGS CAEMDVWEANSISNAFTAHP CDSV SQTMC DG 261

Qy 240 DGCGGTY--SDNRYGGTCDPDGCDWNPYRLGNTSFYGP GSSFTLDTTKKLT VVTQFET-- 295

Db 262 DSCGGTYSASGD RYSGTCDPDGCDYNPYRLGNTDFYGP G--LTVD TNSPFTVVTQFITDD 319

Qy 296---SGA---INRY YVQNGVT FQQPNAELGSYSGNELND DYTAEAEFGGSSFS DK-GGL 348

Db 320 GTSSGTLTEIKRL YVQNGEVIANGASTYSSVNGSSITS AFCESEKTLFGDEN VFDKHGGL 379

Qy 349 TQFKKATSGGMVL VMSLWDDYYANMLWL DSTYPTNETSSTPGAVRGSCSTSSGVP AQVES 408

Db 380 EGMGEAMAKGMVL VLSLWDDYAADMLWLDS DYPVNSSASTPGVARGTCSTDSGVPATVEA 439

Qy 409 QSPNAKVTF SNIKFGPIGSTGNPSGGNPPGGNPPGTTTTRRPATTT-----GSSPG 459

Db 440 ESPNAYVTYSNIKFGPIGSTYSSGSSSGSGSSSSSSSTTTKATSTTLKTTSTTSSGSSST 499

Qy 460 PTQSHYGQCGGIGYSGPTVCASGTT CQVLNPPYYSQCL 496

Db 500 SAAQAYGQCGGQGTGPTTCVSGYTCTYEDAYYSQCL 536

RESULT 10

US-08-676-166A-3

; Sequence 3, Application US/08676166A

; Patent No. 5955270

; GENERAL INFORMATION:

; APPLICANT: Radford, Alan

; APPLICANT: Parish, John H.

; TITLE OF INVENTION: EXPLOITATION OF THE CELLULASE COMPLEX OF

; TITLE OF INVENTION: NEUROSPORA

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David A. Jackson, Esq.

; STREET: 411 Hackensack Ave, Continental Plaza, 4th

; STREET: Floor

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/676,166A

; FILING DATE: 15-JUL-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 1321-1-002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-487-5800

; TELEFAX: 201-343-1684

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 525 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:


```

; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/329,350
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/841,636
; FILING DATE: 30-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,335
; FILING DATE: 17-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,926
; FILING DATE: 04-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,840
; FILING DATE: 28-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/732,181
; FILING DATE: 16-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI96/00550
; FILING DATE: 17-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Shea Jr., Timothy
; REGISTRATION NUMBER: 41,306
; REFERENCE/DOCKET NUMBER: 1716.0510006/MAC/TJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Melanocarpus albomyces
; STRAIN: ALKO4237
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..452
; OTHER INFORMATION: /label= 50K-cellulase-B
US-09-329-350-35

```

```

Query Match          44.6%; Score 1221.5; DB 2; Length 452;
Best Local Similarity 51.8%; Pred. No. 3.5e-87;
Matches 220; Conservative 76; Mismatches 118; Indels 11; Gaps 9;

```

```

Qy      9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANNRWTHATNSSTNCYDGNWSSSTLCPDNECAK 68
| |||||::: | | ||||| | | |||||::: | |
Db     31 ENHPPLTWQRCTAPGNCQTVNAEVVIDANNRWLHDDNMQ-NCYDGNQWTNACSTATDCAE 89

Qy     69 NCCLDGAA-YASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQEFTLLGNEF 126
|::|| | |||:||||::: |||: |||: | || | |::||
Db     90 KCMIEGAGDYLGTYGASTSGDALTLKFVTKHEYGTNVGSRFYLMNGPDKYQMFNLMGNEL 149

Qy    127 SFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQA 186
:||||: | : |||: |||||: | |||: |||: |||||: | | |||||: | :
Db    150 AFDVDLSTVECGINSALYFVAMEEDGGMASYPNQAGARYGTGYCDAQCARDLKFVGGKA 209

```



```

.Best Local Similarity 36.3%; Pred. No. 1.9e-50;
Matches 172; Conservative 71; Mismatches 161; Indels 70; Gaps 17;

Qy      9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNE-C 66
      | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     28 EVHPQITTYRCKTADGCEEKTNIVLDALSHPVHQVDNPFYNGDWGQKPNETACPDLESC 87

Qy     67 AKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKN-VGARLYLMASDTT---YQEFTLL 122
      | : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     88 ARNCIMDPVSDYGRHGVSTDGTSRL---KQLVGGNVVSPRVYLL--DETKERYEMLKLT 142

Qy    123 GNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PRDL 179
      | | | | : | | | : | | | | | | | | | | | | | | | | | | | | |
Db    143 GNEFTFDVDATKLPCGMNSALYLSEMDATGARSE--LNPGGATFGTGYCDAQCYVTP--- 197

Qy    180 KFINGQANVEGWEPSSNNANTGIGGGHSCCSEMDIWEANSISEALTPHPCTTVGQEICEG 239
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | |
Db    198 -FINGLGNIE-----GKGACCNEMDIWEANARAQHIAPHPCSKAGPYLCEG 242

Qy    240 DGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGGPGSSFTLDTTKKLTVVVTQFETSG-- 297
      | : | | : | | | | | | | : | | | : | | | : | | | | | | |
Db    243 AEC-----EFDGVCDKNGCAWNPYRVNVTDYEGEGAEFRVDTTRPFSVVTQFRAGGDA 295

Qy    298 -----AINRYYYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEAEAFGGSSFSKKGGLTQ 350
      : | | : | | : : : | : : | : | | | : | : | : | : | :
Db    296 GGGKLESIYRLFVQDGRVIESYVVDKPLPPTDRMTDEFCAAT----GAARFTELGAMEA 351

Qy    351 FKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQS 410
      | : | | | : | : | : | | | | | | | | | | | | | | : | :
Db    352 MGDALTRGMVLALSIIWWSEGNDMNWLDSE-----GEAGPCDPDEGNPSNIIRVQ 399

Qy    411 PNAKVTFSTNIKFGPIGSTGNPSGGNPPGGNPPGTTTTTRRPATTTGSSPGPTQSH 464
      | : : | | | : : | | | | : | | | | | | | | | | | : | :
Db    400 PDPEVVFSNLRWGEIGST-YESAVDGPVGKKGKKGKAPAAA---GDGNGKEKSN 449

```

```

RESULT 4
S38794
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (Humicola grisea)
N;Alternate names: beta-glucan cellobiohydrolase; exoglucanase
C;Species: Humicola grisea var. thermoidea
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S38794; S08240; A45869
R;Radford, A.
submitted to the EMBL Data Library, June 1991
A;Reference number: S38794
A;Accession: S38794
A;Molecule type: DNA
A;Residues: 1-525 <RAD>
A;Cross-references: UNIPROT:P15828; UNIPARC:UPI000012BE0F; EMBL:X17258; NID:g2760;
PIDN:CAA35159.1; PID:g2761
A;Note: this is a revision to the sequence from reference S08240
R;de Oliveira Azevedo, M.; Radford, A.
Nucleic Acids Res. 18, 668, 1990
A;Title: Sequence of cbh-1 gene of Humicola grisea var. thermoidea.
A;Reference number: S08240; MUID:90175006; PMID:2308855
A;Accession: S08240
A;Molecule type: DNA
A;Residues: 1-299,'H',301-525 <DEO>
A;Cross-references: UNIPARC:UPI00001729F6; EMBL:X17258
A;Note: the authors translated the codon CAG for residue 87 as His
A;Note: this sequence has been revised in reference S38794
R;Azevedo, M.; de, O.; Felipe, M.S.S.; Astolfi-Filho, S.; Radford, A.
J. Gen. Microbiol. 136, 2569-2576, 1990
A;Title: Cloning, sequencing and homologies of the cbh-1 (exoglucanase) gene of Humicola grisea
var. thermoidea.
A;Reference number: A45869; MUID:91178527; PMID:2127803

```

A;Accession: A45869
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-20, 'R', 22-34, 'K', 36-86, 'H', 88-141, 'V', 143-157, 'Y', 159-237, 'QQH', 241-244, 'I', 246-299, 'H', 301-525 <AZE>
 A;Cross-references: UNIPARC:UPI00001729F7; GB:M64588; GB:X17258
 A;Note: this sequence has been revised. See entry S08240
 C;Genetics:
 A;Gene: cbh-1
 A;Introns: 138/1
 C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homology
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation
 F;494-525/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 59.9%; Score 1641.5; DB 1; Length 525;
 Best Local Similarity 57.3%; Pred. No. 1.2e-91;
 Matches 294; Conservative 74; Mismatches 122; Indels 23; Gaps 8;

Qy	1	QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANNRWTHATNSSTNCYDGNWTSSSTL	60
		: : : :: : : : : :::	
Db	19	QQACSLTTERHPSLSWNKCTAGGQCQTVQASITLDSNWRWTHQVSGSTNCYTGKWDTSI	78
Qy	61	CPD-NECAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE	118
		: : : : : : : :	
Db	79	CTDAKSCAQNCCVDGADYTSTYGITTINGDSLKLFVTKGQHSTNVGSRITYLMDGEDKYQT	138
Qy	119	FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD	178
		: : : : :	
Db	139	FELLGNEFTFDVDVSNIGCGLNGALYFVSMDADGGLSRYPGNKAGAKYGTGYCDAQCPRD	198
Qy	179	LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE	238
		: : : : : : : :	
Db	199	IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE	258
Qy	239	GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVVTQFETS--	296
		: : : : :	
Db	259	GDSCGGTYSNERYAGVCDPDGCDFNYSYRQGNKTFYKG--MTVD'TKKITVVVTQFLKDAN	316
Qy	297	---GAINRYYYQNGVTFQQPNAELGSYSGNELNDYCTAEAEFGG-SSFSDKGGLTQFK	352
		: : : : : : : : : :	
Db	317	GDLGEIKRFYVQDGKIIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG	376
Qy	353	KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSSTPGAVRGSCSTSSGVPAQVESQSPN	412
		: : : : : : : : : :	
Db	377	KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN	435
Qy	413	AKVTFSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQS	463
		: : : : :	
Db	436	SNVVFNSNIRFGPIGSTVAGLPAGNGGNGGNPP---PPTTTTSSAPATTTTASAGPKAG	492
Qy	464	HYGQCGGIGYSGPTVCASGTTTCQVLNPHYYSQCL	496
		: : : :	
Db	493	RWQQCGGIGFTGPTQCEEPYICTKLNWYSQCL	525